SPEECH RESEARCH '89

INTERNATIONAL CONFERENCE
JUNE 1–3, 1989, BUDAPEST

PROCEEDINGS

AZ MTA NYELVTUDOMÁNYI
INTÉZETE, BUDAPEST 1989

MAGYAR FONETIKAI FÜZETEK
Hungarian Papers in Phonetics
21.

PROCEEDINGS
of the
SPEECH RESEARCH '89
International Conference,
June 1–3, 1989, Budapest

Edited by

TAMÁS SZENDE

LINGUISTICS INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST 1989

HU ISSN 0134--1545

Az MTA Nyelvtudományi Intézete, Budapest 1989

Felelős kiadó: HERMAN József, az MTA Nyelvtudományi Intézetének igazgatója
Készült 400 példányban, 25 (A/5) ív terjedelemben.

Hozott anyagból sokszorosítva.
8918730 MTA Sokszorosító, Budapest. F.v.: dr. Héczey Lászlóné

CONTENTS

SPEECH AND HEARING DISORDERS

PERCEPTION

SPEECH PROCESSING

PREFACE

Half-way between Tallinn and Aix-en-Provence, geographically as well as timewise, the SPEECH RESEARCH '89 international conference is intended to serve a double purpose. First, now that new scientific achievements spring up almost every day in one research center or another, and it is becoming increasingly difficult to keep abreast of progress in our respective disciplines, the organizers of this conference wanted to afford the possibility for experts in the various branches of speech research to gather an extra time between two world congresses and present their most recent results to, and develop or maintain personal contacts with, one another. Second, we also wanted to create an occasion for speech researchers from all parts of the world to get acquainted, in situ, with the directions and accomplishments of speech research as it is practised in Hungary. It is our hope that this encounter will further increase the chances of Hungarian colleagues and research teams to join the mainstream of international speech science.

The present volume contains a generous selection of the papers submitted to the conference, thematically arranged. The topics of most papers could have been classified under several section headings: in a number of cases the organizing committee had to make arbitrary decisions about the particular sections such papers were assigned to.

We want to express our thanks to all participants of the conference for having accepted our invitation, and especially to all those people who have helped, in one way or another, to make this conference possible and successful.

The Organizing Committee

WIGNER DISTRIBUTION ANALYSIS OF STOP CONSONANT RELEASE TRANSIENTS LABIALS, VELARS AND LABIO-VELARS.

Grzegorz Dogil, Universität Bielefeld, Fakultät für Linguistik.

Wolfgang Wokurek, Universität Wien, Institut für Nachrichtentechnik.

Introduction

It has been argued that the release burst provides the most direct acoustic manifestation of stop consonant place of articulation. However, of the three acoustic segments making up the burst - i.e. the release transient, the fricative segment, and the aspirative segment - only the first, the transient, has a coherent source. Whereas the fricative and aspirative segments have the 'stochastic noise source', the spectrum of the transient should contain parameters characteristic of the vocal tract resonances immediately after the release. Thus the acoustic characteristics of transients should, in theory, contain properties associated with different places of articulation in purer form than do the bursts. However, standard methods of acoustic time-frequency analysis lack the required temporal resolution necessary for the analysis of transients.

The analysis tool: Wigner Distribution

It is well known that joint time-frequency analysis of speech signals can be performed using wide-band or narrow-band spectrograms. A wide-band spectrogram analyzes a signal with good time resolution but poor frequency resolution, thus showing the periodic time structure of vowels but broadening their formants. The narrow-band spectrogram, on the other hand, offers good frequency resolution but poor time resolution. This behaviour, i.e., the feasibility of good resolution in only one direction is a specific feature of the spectrogram.

There exists, however, a joint time-frequency representation known as Wigner Distribution (Claasen & Mecklenbräuker 1980) that overcomes the difficulty of not being able to adjust both good time resolution and frequency resolution simultaneously. Although Wigner Distribution has ideal resolutions, it is not applicable to the analysis of speech signals without modification. A suitably modified version, the Smoothed Pseudo Wigner Distribution (abbr.: SPWD) was introduced by Flandrin & Martin (1983) and has been applied to the analysis of speech signals by Wokurek et. al. (1987) and Dogil (1988).

SPWD allows the user to achieve exactly the time resolution and the frequency resolution required for a specific analysis task. In particular, by adjusting the SPWD resolutions to that of any specific spectrogram, the SPWD will simulate this spectrogram. More important, however, is the possibility of achieving better resolutions than those obtainable with spectrograms.

In this paper, we employ high-resolution SPWD to analyze speech signals, in particular, to display transients and noise-bursts. Due to its improved resolutions, SPWD here yields a time-frequency representation in which the finer signal structures are considerably more evident than in spectrograms.

A transient is shown in SPWD as a combination of a vertical line (parallel to the frequency axis) followed by a number of horizontal lines (parallel to the time axis). The vertical line represents an impulse-like signal component caused by a sudden release of air compressed behind a constriction as e.g. in stop consonants. The horizontal lines are interpreted as damped oscillations (forment frequencies) forming the response of the vocal tract to the plosion-pulse.

Noise bursts will be represented in SPWD as an irregular grid, filling a specific area of the time frequency plane. In addition to the bandwidth and the duration of the noise-burst, the special shape of the region displays time-variant features.

Material

The Baule[1] stops [k] [p] [g] [b] [kp] [gb] were placed in an intervocalic position and combined with the vowel [a]. As Baule is a tone language containing 3 level tones (high, low and mid), the stop [kp] was additionally placed into varying tonological contexts (mid-high, high-low, low-high). These 8 VCV collocations were spoken in isolation by Kwamé Akpetou at regular intervals of about 2 seconds. The recordings were made in a sound studio in the University of Bielefeld.

---

[1] Baule is one of the Akan languages of the Kwa group. It is one of the major native languages of the Ivory Coast.

Analysis

The analog signals have been digitalized with the 20 kHz sampling rate, and the 50 ms of the signal corresponding to the release burst and the first periodicities of the vowel have been cut out.

Labio-velars

In the production of the labio-velar stops the lips and the back of the tongue are involved in a complex articulatory action which involves the closure and the simultaneous release of the air-stream by both articulators. With the standard techniques it was difficult to measure how exact this complex articulatory gesture was. Figure 1 illustrates the initial 50 ms of [kp] followed by [a]. Note the short high-frequency disturbance of the waveform near to point 360 on the time scale. Although there is apparently no plosion which could be identified in the waveform (except maybe the ominous perturbance at point 360) the Wigner Distribution analysis of this signal shows a very clear pattern.[2] Consider figure 2 below. The vertical scale is the frequency scale (10 kHz), the horizontal scale is the time scale (50 ms). Intensity (50 db dynamics) is represented by contour lines, equally spaced in 1dB steps. The release transient is clearly represented by two vertical islands, which stretch above 2500 Hz and below 7500 Hz on the frequency scale, and a very clear transient, which lies at approximately 3200 Hz on the frequency scale and is approximately 10 ms long. Note that there is apparently no fricative segment and no aspirative segment following the release of [kp].

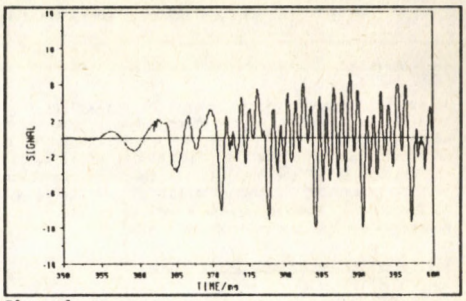


Figure 1.

Sometimes it is the case that the low intensity of the fricative noise following the transient is fully masked by the naturally more intense vowel. SPWD is delicate enough to check this exactly by taking a closer look at the pre-vocalic part alone. For this purpose only the first 10 ms of the burst have been edited from the signal. The SPWD analysis of this part of the signal is given in figure 3 below. The most salient part of the release of the labio-velar stop is clearly the transient with high energy at 3200 Hz. The part of the release transient which was visible as two small vertical islands in figure 2, is now explicitly represented through a clear vertical pattern of low intensity occupying the whole frequency area above 2500 Hz. There are no signs of either friction or aspiration in this representation of stop release.

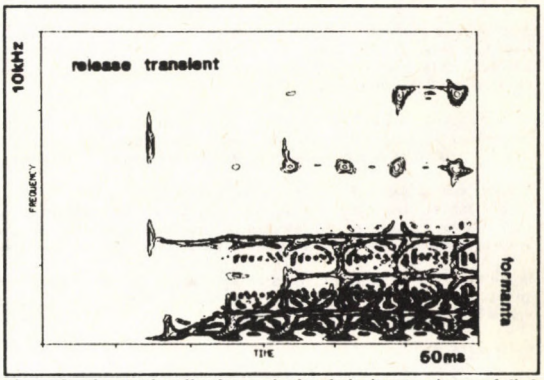


Figure 2. Wigner Distribution analysis of the burst release of [kp] and of the first vowel periodicities of [a] in [akpa].

Velars

We will now compare the characteristics of the labio-velar plosion with the characteristics of the other two competitive 'simple' places of articulation. Due to lack of space we will present only the decisive 10 ms following the plosion. Figure 4 illustrates the SPWD pattern of the pre-emphasized signal. The most conspicuous pattern in figure 4 is the release transient at 1200 Hz. The rest of the release transient has rather weak intensity and its representation (the vertical island in the high-frequency region) shows a much shorter time resolution than the release of the labio-velar [kp] (cf. figure 2).

---

[2] High-frequencies have been pre-emphasized for the purpose of the analysis, and 50 db dynamics has been chosen.

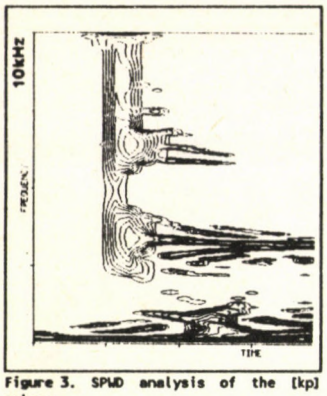


Figure 3. SPWD analysis of the [kp] release.

Another difference to the representation of the labio-velar release is the presence of high frequency noise immediately following the plosion of [k], and the presence of some low-intensity, high-frequency noise preceding the transition to the vowel. These noise portions of the signal must, probably, be associated with the fricative and the aspirative segments of the burst phase.

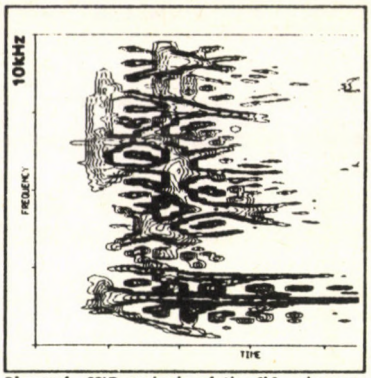


Figure 4. SPWD analysis of the [k] release.

Labials

In order to have a SPWD snapshot at the release of the labial articulation we cut out a part of the signal (10 ms) that immediately precedes vowel periodicities. Again high-frequencies have been pre-emphasized prior to analysis. The results of SPWD analysis of this part of the signal are given in figure 5.

The interpretation of the SPWD patterns of figure 5 is not as easy as in the cases of labio-velar and velar releases. The response of the vocal tract to the labial release seems to be very weak as there are no vertical structures (vertical islands) which so clearly illustrate the release phase in terms of SPWD for other stops. The transient (at approximately 1500 Hz), which is the only sign of the plosion, is very weak. The longitudinal energy concentrations at 3000 and 6000 Hz may hardly be described as transients. What is clearly noticeable is the concentration of energy in the lowest frequency regions of the burst. This low-frequency energy has been found to be one of the distinctive features of the labial burst in the standard literature (cf. Borden & Harris 1982, 181ff.). SPWD time-frequency resolution allows us to make a more delicate observation in this case too. The strong and highly concentrated low-frequency energy of the labial burst is apparently preceded by slighter and more broadly distributed energy in higher frequencies (above 3000 Hz). The highest frequencies of the initial phase of the labial also show light fricative energy.

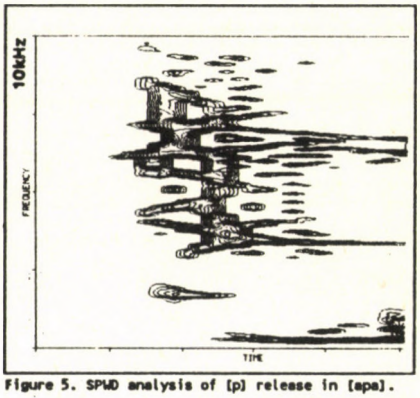


Figure 5. SPWD analysis of [p] release in [apa].

Discussion of the SPWD analysis of stop release

The delicacy of the Wigner Distribution analysis in the time--frequency area allows us to make a number of interesting observations about the characteristics of stop consonants. SPWD is particularly useful, and actually currently unique, as a method of analyzing the weakest and shortest acoustic segments of a stop - i.e. the release transient. As figures 3, 4 and 5 show the release is most explicit in labio-velars (fig. 3), it is weaker in velars (fig. 4), and in labials (fig. 5) it can be hardly noticed at all. Labio-velars also have the strongest transient (vocal tract response to stop release), which is situated at approximately 3200 Hz. The transient of the velars is also strong but it is much lower in frequency (1200 Hz). The transient of the labial is very weak and it occupies the frequency of approximately 1500 Hz. The velar has the strongest fricative segment, which is also present (but weaker) in the labial articulation, and seems to be missing totally in the labio-velar.[3] The overall length of the burst phase is almost the same in all consonants. Hence the

[3]The SPWD analysis of a frequency band above the transient did not show any fricative energy in this region.

distinguishing cues seem to be the 'strength' and the position of the release transient.

Other characteristic patterns of stop consonant place of articulation which have been observed with different measurement methods can be seen in the SPWD patterns too (cf. Dogil, 1988). However, it is the additional information about the transient and other weak energy parts of the signal where the SPWD analysis shows clear advantages over other measurement techniques. The question to be asked here is whether this information about the minute parts of the signal actually has any importance for invariance in speech. The answer must be 'yes'. Acoustic theory tells us that the release transient provides essentially an acoustic snapshot of the vocal tract, and the information about the place of articulation that it reveals is purer than the information which can be detected from the whole burst. Also, perceptual experiments (cf. Repp & Lin 1988) show that listeners can identify consonants and vowels from isolated transients with no less difficulty than they can identify them from whole release bursts. Actually such perceptual experiments may be first performed given the SPWD analysis. In the experiments of Repp & Lin which we were referring to, the human speaker was producing isolated transients by 'mouthing' /b,d,g/ in various vocalic contexts. Not questioning the skills of the natural talker in performing such a complex task it must be pointed out that the produced transients can hardly be called natural. The method of SPWD analysis, which enables us to isolate the natural transient opens new possibilities for acoustic phonetic experimentation. The transient can be isolated form the natural signal, thus it can be provided as a test material. Because it can be isolated, it can be manipulated and checked in the experimental situation again.

The material discussed in this study does not allow us to make strong claims about the invariant cues to the labio-velar place of articulation. The study has to be extended to other languages,[4] and to other contexts. Important seem to be these cues which can be found in the micro-phonetic area and which were depicted by SPWD analysis. The transient of the labio-velar stop seems to be stronger than the transients of labial and velar consonants and it is also higher in frequency. Most interesting is the 'strong' release of the labio-velar, which is illustrated as the broad vertical island in figure 4. This is actually the only cue which gives some reason for the description of this plosive as a 'complex' one. In any case, SPWD analysis clearly shows that labio-velars have a single release.

References

Anderson, S. (1976). 'On the description of multiply articulated consonants,' Journal of Phonetics 4, 17-27.

Borden, G. & Harris, K. (1982/1984). Speech Science Primer, Williams & Wilkins, London.

Claasen, T. & Mecklenbräuker, W. (1980). 'The Wigner Distribution - a Tool for Time-Frequency Signal Analysis', Philips J. Res. 35, pp. 217-250, 276-300, 372-389.

Dogil, G. (1988). 'On the acoustics of multiply articulated stop consonants (labio-velars),' Wiener Linguistische Gazette, 42/43, 3-56.

Flandrin, P. & Martin, W. (1983). Pseudo Wigner Estimators for the Analysis of Nonstationary Processes. Proceedings of the IEEE ASSP Spectral Estimation Workshop II, pp. 181-185. Tampa, Florida.

Ohala, J. (1979) 'Universals of labial velars and de Saussure's chess analogy,' Proceedings of the Ninth International Congress of Phonetic Sciences, vol 2, 41-47, Copenhagen, Institute of Phonetics.

Ohala, J. (1985) 'Around FLAT,' in V. Fromkin (ed.) Phonetic Linguistics, 223-241.

Repp, B., & Lin, Hwei-Bing. (1988). 'Acoustic properties and perception of stop consonant release transients,' ms. Haskins Laboratories.

Wokurek, W., Kubin, G., & Hlawatsch, F. (1987). 'Wigner distribution - a new method for high-resolution time-frequency analysis of speech signals,' in Proceedings of the 11th International Congress of Phonetic Sciences, Academy of Sciences of the Estonian S.S.R., Tallinn, Vol. 1, 44-47.

Wokurek, W., Hlawatsch, F. & Kubin, G. (1987). 'Wigner Distribution Analysis of Speech Signals', Proceedings of the International Conference on Digital Signal Processing, Florence, Italy, Sept. 7-10, 1987.

[4] Cf. Anderson (1976, 22-23) and Ohala (1979, 1985).

PERIODICITY-NOISE ACOUSTIC SPACE FOR SPANISH CONSONANTS

Jorge GURLEKIAN, Horacio FRANCO, Miguel SANTAGADA
Laboratorio de Investigaciones Sensoriales
CONICET-UBA, CC 53, 1453, Buenos Aires, Argentina

Introduction

The voiced-unvoiced distinction has been used in several recognition systems as a preliminary classification of speech sounds in order to use this information to make final decisions or to direct further processing more selectively. As this distinction is only based upon the presence or absence of a low frequency periodicity in the sound signal the question which arises is if those sounds which 1) combine periodicity and noise 2) have a different degree of periodicity or 3) have different noise components could define new categories according to the analysis of the periodicity-noise levels involved in their production.

Besides, supporting this idea, some experiments on psychoacoustic categories such as roughness (1) have shown that the subject numerical estimations for the set of spanish consonants conform a continuum along which the sound categories are distributed which indicates a continuous variation of periodicity-noise levels.

Herein, we examine the relevant acoustic characteristics of Spanish consonants at intervocalic positions aiming to determine an acoustical space where different types of consonants could be represented according to periodicity-noise levels and to test a new method designed for this purpose. Several systems of Voiced-Unvoiced detection are based on the use of the short time autocorrelation function (AC), either of the speech signal or of some transformation of it. Most of these methods rely on the detection of the highest peak of the autocorrelation function - within a certain range of lag values which correspond to the expected range of pitch values- and to compare it with a defined threshold. In this way a speech segment is considered as unvoiced if its value is below the threshold. Many different systems of V/U detection are derived of this method but esentially they are all similar.

A different approach to this problem has been explored (3) which is based on the discrimination of the characteristic pattern of the autocorrelation function of the residual signal. In this way from the observation of different speech utterances the pattern of the AC function shows clear differences between the voiced and unvoiced segments which can be distinguished using a pattern recognition method. In this paper the concept of the general characteristic of the AC function is applied but to the emphasized speech signal and using a different feature of the AC function.

Methods

For the purpose of making measurements that allow us to quantify the periodicity degree and the noise level present in the consonants under study an experimental evaluation using a correlation coefficient was performed between adjacent values of the short time autocorrelation function over a predetermined delay range. Zero crossing rate estimations to evaluate the noise contribution and estimations of the pitch and voicing based on the measurement of the cepstrum peak in the same selected frames were made for comparison with the proposed coefficient.

The Correlation Coefficient

For the periodic sounds the shape of the AC function is also periodic with a maximum in lag=0 and others relative maxima with similar amplitude located at multiples of the fundamental frequency. Between these peaks some other peaks can appear which are due to the impulse response of the vocal tract. As these peaks can have also high amplitudes the approach of peaking the maximum of the AC function is not enterely reliable. Another characteristic of the AC function -which is used herein- is that the correlation between adjacents values of the AC is generally high even for lag values around the pitch period.

For the unvoiced sounds the AC function has also a maximum in lag=0 and then its amplitude rapidly decreses as the lag values increse until a level around zero is reached where the shape is completly unestable and noisy. Moreover the correlation between adjacent values for lag values between 5 and 20 msec is very low and sometimes negative.

A correlation coefficient for adjacent values of the AC function for lag values within 5 and 20 msec has been used. Periodic sounds should reveal positive values while aperiodic ones should present small or even negative values.

Formally it will be defined as follows:
Given s(n) the speech signal and w(n) the Hamming window of lenght L from which we obtain the finite signal w(n) = s(n).w(n)

Then the normalized and unbiased AC is obtained as:

$$AC(j) = \frac{\sum_{n=0}^{L-j-1} x(n).x(n+j)}{VAR \cdot (L-j)}$$

where VAR is:

$$VAR = \frac{1}{L} \sum_{n=0}^{L-1} x(n)^{**2}$$

now we define ACC the Autocorrelation Coefficient within the LAG range (index j) determined by LAGmin and LAGmax

The denominator is a normalization factor to obtain amplitude values between -1 and +1. So, for periodic sounds ACC should be between 0 and +1 and 0 and -1 for non periodic sounds.

$$ACC = \frac{\sum_{j=LAGmin}^{LAGmax-1} AC(j) \cdot AC(j+1)}{\sum_{j=LAGmin}^{LAGmax-1} AC(j)^{**2}}$$

Material

The corpus consisted of the sequence of the Spanish Consonants / β ð γ p t k s f x ʧ ζ r R m n ɲ l/ in combination with the vowels /a i u / uttered in VCV sequences by 10 male speakers native from Buenos Aires. Our measurements were made not only on determinated segments of the sounds in study, but also on segments of the adjacent vowels taken herein as reference. The recorded material was filtered by a low-pass eliptic filter, digitalized to 10000 samples by second, and stored in disk. The stored waveform was emphasized in the high frecuencies and filtered in its continuous component before the short time analysis was performed. The analysis of the whole emission was made, using a Hamming window of 25 ms, whose displacement for the computation takes place every 10 ms covering the complete secuence. In order to establish the statistical characterization, five representative points were selected, two of them corresponds to the major spectral stability in the preceeding and following vowel respectively. This zone was defined as one where the spectral derivative was minimal and the energy remains high; a third point is the medial one of the consonantical segment, measured in the medial frame between the first formant transitions of the adjacent vowels. For the consonants /p t k t / a careful marking was performed to consider the frame associated with the short noises. Two other points were selected at the maximum overall amplitude derivatives adjacent to the consonant.

Results

The results for ACC means shown in Table 1 allowed to separate the consonants for each of the context vowels in two clear pair of sets, one composed by the sounds /s ʧ ζ f p t k/ with negative ACC's and a second set formed by /m n ɲ l r R β ð γ/ with positive values of ACC. Consonant /x/ shares the first group for the /i-i/ context and the second group for the /a-a/ and /u-u/ context holding the minimum values in those cases. It must be pointed out that the mean values of ACC measured at the preceding vowel /i/ were the lowest -even negative- of the set of vowels, while the preceding and following /u/ hold the highest. The coefficients were in average 0.60 for /a/, 0.90 and 0.20 for /i/. Standard deviations showed lowest values for the consonants at the extreme of the range around 0.07 but in the medial range this value reached in average 0.30.

Table 1.

C	ACC /a/		ACC /u/		ACC /i/		Cepstrum /a/		Cepstrum /u/		Cepstrum /i/		Zero Crossings /a/		Zero Crossings /u/		Zero Crossings /i/	
β	.88	γ	.96	η	.92	n	.32	l	.30	n	.23	β	1164	β	862	m	1074
η	.56	β	.93	m	.77	m	.28	m	.28	m	.19	η	1329	r	866	η	1145
m	.55	r	.92	β	.62	l	.26	n	.26	η	.16	n	1623	γ	905	n	1247
r	.54	η	.90	n	.61	η	.23	η	.24	l	.15	r	1639	η	1082	β	1454
δ	.52	R	.83	R	.49	δ	.19	γ	.13	γ	.12	R	1650	R	1262	r	1788
γ	.45	l	.73	δ	.49	β	.17	β	.12	δ	.11	m	1749	n	1513	R	1972
n	.44	n	.48	r	.35	R	.14	R	.10	β	.12	l	1752	l	1529	δ	2054
l	.39	δ	.46	γ	.29	r	.12	δ	.10	ζ	.09	δ	1800	m	1756	l	2098
R	.35	m	.43	l	.17	γ	.11	r	.08	R	.09	γ	1898	x	2121	p	2262
x	.21	x	.21	p	-.08	t	.08	x	.08	x	.08	p	2156	k	2121	γ	2419
k	-.10	t	-.11	f	-.13	s	.08	tʃ	.08	s	.08	k	2180	δ	2172	t	2466
t	-.27	tʃ	-.15	t	-.23	p	.07	f	.08	tʃ	.08	x	2227	t	2294	k	2913
p	-.33	k	-.15	x	-.37	f	.07	s	.07	r	.08	t	2278	p	2447	f	2933
f	-.39	f	-.16	k	-.48	tʃ	.06	p	.07	p	.07	f	3061	f	2690	x	3078
s	-.78	p	-.25	ζ	-.81	k	.06	ζ	.07	f	.07	tʃ	3603	tʃ	2784	tʃ	3509
tʃ	-.79	ζ	-.48	tʃ	-.82	ζ	.06	t	.06	k	.07	s	3956	ζ	3305	s	3800
ζ	-.85	s	-.50	s	-.83	x	.06	k	.06	t	.06	ζ	4007	s	3313	ζ	3850

Table 1: Mean values of the Autocorrelation coefficient, Cepstrum and Zero Crossings for each vocalic context ordered for the most periodic sounds to the least ones.

Conclusions

The ACC measurements confirm the former results that the consonants / β δ γ / -periodics with low amplitude- present periodicity noise levels closed to periodics with medium amplitude sounds such as /l m n /. On the other hand, the measurement of the voiced fricative /ζ/ has shown noise levels closed to those found in the unvoiced fricative /s/, which indicates a clear distinction among / β δ γ/ and /ζ/. For the Cepstral Peak Amplitude estimations and Zero Crossing rate estimations the values obtained reflect a similar tendency as the described for the ACC estimations. Periodic sounds can be discriminated from noisy ones but the definition of new categories within them would require a finest statistics.

References

1. GUIRAO, M.: Toward a Psychoacoustical Classification of Speech Sounds, Transactions of the Committee on Speech Research/Hearing Research. The Acoustical Society of Japan, S81-34, 1981, 267--274.
2. SANTAGADA, M.--GURLEKIAN, J.A.: Spanish Voiced stops in VCV contexts: are they fricative variants or approximants ?. Abstracts of the 1st Int. Conference on Experimental Phonostylistics & Sociophonology and SPeech Acoustic Variability ,UFSC. Florianopolis. Brasil 1988, 99.
3. YANAGIDA, M.: Personal communication.

INTERNATIONAL RECOMMENDATIONS ON SPEECH QUALITY

György LAJTHA
Central Administration of the Hungarian
Posts and Telecommunications, Budapest

1. Introduction

The International Telegraph and Telephone Consultative Committee (CCITT) has played a very important role in drafting recommendations for defining the necessary speech quality which enables conversations also between subscribers of different mother tongues. In view of this, when planning the intelligibilty of a connection, it has to be made sure, that intelligibility is not only specified for sentences or words but also for logatoms, in other words, so far as comprehension is concerned one must not rely on human brain's ability of interpretation. In analog environment (transmission and switching) the definition of speech quality was derived from the users' opinion. The opinion score reflects in general several impairments in a single value, but it is not simple to find correlation between the mean opinion score (MOS) and the technical performances of the transmission path. A second level of the subjective quality assessment is the intelligibility, from which the articulation factor can be derived. The connection between the subjective and objective measures for analog transmission was developed by Fletcher and Richards. Their results were the basis of the recommendations in which the necessary quality is defined by simple objectively measurable characteristics, as level, noise, linear and non-linear distortion.

In recent years several new methods and devices have appeared, where the earlier introduced objective measures cannot charactarize in simple manner the opinion of the users. On the first place we have to mention the digital transmission and switching. The CCITT has standardized several digital speech coding processes (e.g., A- and μ-law PCM and narrow- and wideband ADPCM, respectively) and recommendations for other processes, such as DCME (digital circuit multiplexing equipment). Further there are other digital speech processing instruments and the artificial speech is playing also an increasing role in the enquiry services and the machine-man connections. So it was necessary to find new methods and rules which can help to find correlation between subjective opinion and objective maesurements, and to create recommendations which give planning rules and the values in the recommendations should be measured in a short time. They should give results reproducable with good accuracy. Now I try to give an overview of the results of the subjective and objective quality evaluation techniques achieved by the CCITT.

2. Methods for subjective assessment

Subjective assessment methods can be categorized in numerous ways: laboratory versus field experiments, conversational versus listening tests, and single stimulus rating versus paired comparisons. Results can be obtained in terms of overall quality, listening effort, percentage difficulty, intelligibility, naturalness or some other scale. Laboratory experiments typically provide results faster and less expensively than field trials, but may be difficult to design so that sufficient realism is provided. Field trials can offer the necessary realism but are costly to conduct, and it is often difficult to maintain control over all the variables that may influence customer opinion.

Several conversation test methods, and post-call interviews are recommended, but it is quite difficult to evaluate their results. On the other hand, they do not depend on the applied speech processing technique. They are similar to the opinion tests.

Scrutinizing the quality by opinion score a large number of listeners are qualifying the speech by a 5 stage scale. The meanings of the stages are the following:

Table 1
Subjective assessment scales

quality	listening effort	degradation
excellent	complete relaxation possible, no effort required	degradation is inaudible
good	attention necessary, no appreciable effort required	degradation is audible but not annoying
fair	moderate effort required	degradation is slightly annoying
poor	considerable effort required	degradation is annoying
bad	no meaning understood with any feasible effort	degradation is very annoying

For synthetic speech a detailed quality rating has been developed, which can be seen in the following table.

Table 2

PRONUNCIATION	STRESS	SPEED	INTELLIGIBILITY	DISTINCTNESS	COMPREHENSIBILITY	NATURALNESS	PLEASANTNESS
Incorrect pronunciation is:	Incorrect stress is:	The syllable rate is:	Understanding word by word is:	The Speech is:	Understanding the message is:	The voice sounds:	The voice sounds:
1 Very annoying 2 Annoying 3 Rather annoying 4 Slightly annoying 5 Not annoying	1 Very annoying 2 Annoying 3 Rather annoying 4 Slightly annoying 5 Not annoying	1 Much too fast 2 Too fast 3 Somewhat too fast 4 Somewhat too slow 5 Too slow 6 Much too slow	1 Very hard 2 Hard 3 Rather hard 4 Rather easy 5 Easy 6 Very easy	1 Very slurred 2 Slurred 3 Rather slurred 4 Rather clear 5 Clear 6 Very clear	1 Very hard 2 Hard 3 Rather hard 4 Rather easy 5 Easy 6 Very easy	1 Very unnatural 2 Unnatural 3 Rather unnatural 4 Rather natural 5 Natural 6 Very natural	1 Very unpleasant 2 Unpleasant 3 Rather unpleasant 4 Rather pleasant 5 Pleasant 6 Very pleasant

It gives more information about the improvement of the quality. The average of the different opinion results is the MOS.

A special method fitting better to digital speech processing is based on the Modulated Noise Reference Unit (MNRU). The MNRU produces random noise with amplitude proportional to the instantaneous speech amplitude (speech-correlated or multiplicative noise). This noise is perceptually very similar to the quantization noise produced by many speech coders, especially those employing non-linear companding laws

and operating generally at 16 kbit/s or more.

The ration of speech level and multiplicative noise level expressed in dB is called Q. For a given coder, an equivalent Q value can be determined by means of the listening test methods. The Q concept has been incorporated into the transmission rating models described in several CCITT recommendations.

Further possible methods are multiple paired comparisons between all systems under test and all reference conditions, and articulation tests of MNRU conditions and digital systems.

The different subjective methods can be subdivided in three classes: the absolute category rating (ACR), the degradation category rating (DCR), and the equality threshold method. The most suitable of these can be chosen for a given purpose.

3. Methods for objective assessment

Methods for objective assessment, in general, attempt to provide a functional relationship, $Y_S = f(X_1, X_2, \ldots, X_n)$, between a measure of customer opinion along a subjective scale, Y_S and levels of impairments, X_n. Here the primary aim is to harmonize the subjective qualification with objective measurements that can quickly be performed.

Three classes of methods have been developed. Each of them tries to describe the quality of the transmitted speech by a single value. The first class derives it by a transformation $R = \tau(S_{in}, S_{out})$. Such transformations are the modified version of correlation function, cepstrum distance mehod or energy distribution. The disadvantage of these is that some constant must be chosen to have a good correlation between R and MOS.

Using cross correlation we find R represents the speech quality only in the domain 0,98 ... 1,00. On the other hand, it is quite difficult to expand this domain to get detailed information. The cepstrum distance (CD) can be calculated using Fourier transform. The number of samples, the weighting of speech level, the method of overlapping of segments has a great influence on the results. The developed methods give MOS $\pm$ 0,25 for the same speech quality, giving uncertainty of half the class, making international comparison impossible. There are similar difficulties with the energy comparison where the choice of limits of frequency bands, the weighting of them influence the result. Up to now there is no definite suggestion for such a transformation. The research continues in this field.

The second class is based on a computer program which gives the measure of speech quality computed from the primary transmission parameters, such as level, distortions, noise, etc. AT&T and the French PTT developed such systems. Enhancing the number of parameters, they give quite good and reproducable results for analog transmission and also in digital

systems if there is no speech processing used. The two systems are closely related. The disadvantage of the systems is that they cannot be used for synthetic speech and they are not capable of taking stochastic impairments into account.

The third class is the most promising. It analyses the received speech by segments and it is scrutinizing the degradation of the segments comparing them to the sending side or to ideal speech. The segmentation and evaluation need a quite sophisticated computer program. The system is developed by György Takács (Hungary) and he has achieved results independent of the type of impairment and language.

4. Conclusion

The choice of the appropriate method or the proper combination of the methods has not been made uniform yet. The research conducted on several fields gives results where the difference in the results is diminishing, so we can hope that in the next 4 years of the CCITT study period new recommendations will be drafted unifying the methods of measuring the speech quality.

5. References

1. DVORAK, C.--ROSENBERGER, J.: Deriving a subjective testing methodology for digital circuit multiplication and packetized voice systems. IEEE Journal on Selected Areas in Communications, Vol.6. (February 1988).
2. CCITT: Subjective testing methodology for the evaluation of low bit-rate codecs for mobile radio. European Joint Experts Group on low-bit-rate coding for mobile radio. Document COM XII-68. (1985-1988).
3. CCITT: OPINE (Overall performance index model for network evaluation). NTT, Supplement No.3 (part D). Document AP IX-6.
4. CCITT: Calculation of transmission performance from objective measurements by the information index method. France, Supplement No.3 (part C). Document AP IX-6.
5. CCITT: Objective speech quality estimation of non-linear telecommunication devices by LPC cepstrum distance measure. NTT, Document XII-86. (1985-1988).
6. DECINA, M.--MODENA, G.: CCITT standards on digital speech processing. IEEE Journal on selected areas in communications. Vol.6. (February 1988).
7. BIGI, F.--DECINA, M.: CCITT standardization work on speech processing. Telecommunication Journal. Vol.55. XI/1988.
8. ROSENBERGER, J.R.: Quality assessment methods for speech coding. Telecommunication Journal. Vo.55. XII/1988.

THE DURATION AND F-PATTERN OF SVANIAN STOP CONSONANTS

Ivane LEZHAVA
Laboratory of Experimental Phonetics
Tbilisi State University, Tbilisi, Georgia, USSR

Introduction

The phase duration of stop sounds has been investigated on the basis of different languages, but a bulk of problems remains unexplained and undefined.

Svanian stop consonants create a triple system: voiced consonants (b, d, ʒ, ǯ, g), aspirated consonants (p, t, c, č, k, q) and glottalized (ejective) consonants (p̣, ṭ, c̣, č̱, ḳ, q̣).

Methods

Svanian stop consonants are studied by means of spectrography and oscillography methods. The experimental date is built of separate Svanian words pronounced by seven male announcers.

As the initial voiced stops' closure is devoiced, the analyses of closure duration are carried out in the intervocalic and final postvocalic positions.

Results and Discussion

The observation on the closure phase duration of Svanian stop consonants reveals the following:

In each localization series of the plosives the glottalized consonants are distinguished by the greater duration of closure. In the intervocal position the average duration of the uvular consonants exceeds the average duration of the labial consonant closure.

Labial consonants in the final position are characterized by even greater closure duration.

Stop consonants of extreme (peripherial) localization series (labial and uvular) are characterized by maximal closure duration. As for the stops of middle (central) localization series (front and back sibilants), their closure duration is minimal. Dental and velar stops occupy the transitive position between the abovementioned consonants according to duration. Dentals and velars do not differ noticeably by closure duration. This relation can be stated in the following way: among the plosives with maximal closure duration the consonants of the front series are singled out and among the affricates - the consonants of the back series.

The differences among the closure phase durations of the stop consonants are satisfactorily explained by the pecularities of sound articulation.

The longer duration of labial stops' closure phase is explained by the relative autonomy of lip movement from tongue articulation and the possibility of articulatory fore-stallement characteristic of lips.

As for the uvular q, q̇ sounds, their greater closure duration is most probably caused by the necessity to achieve closure by inert organs on comparatively wide area.

As for the glottalized plosives, it is fairly well known that they are characterized by tensed articulation and wide area of articulation closure compared to that of voiced and aspirated consonants. Consequently, if closure strength means greater duration, it is not completely unexpected that their closure duration should be correspondingly longer.

The observation on <u>the duration of noise</u> segments (including aspiration) of stop consonants in the initial prevocalic and intervocalic positions reveals following:

a) In each localization series with longer duration of noise the aspirated consonant is singled out. Among sibilant affricates the noise duration of the glottalized consonants is comparatively longer than that of the voiced consonant. On the other hand, the comparison of voiced and glottalized consonants for the plosives proves of considerable difficulty.

b) In one and the same series of the plosives the sounds of front localization are characterized by the shortest duration of noise: labial < dental < velar. The duration difference between the dental and velar plosives can be determined by the flexibility of the tongue. On the other hand the shorter duration of labial plosives can be explained by the fact that, besides flexibility, the lips are characterized by autonomy. At the same time the area of front resonator obstruction can bear considerable importance.

c) The study of the affricates has revealed the following: the sounds of the front localization have the longest noise, thus, whistling sibilant > hushing sibilant > uvular affricate. The articulation flexibility is the cause of durational diversity of affricates as compared to the rest: the differences among affricate durations can probably be determined by the size of articulatory obstruction (noise-producing focus) that, on its own part, depends on the flexibility of articulation and the place of obstruction.

In order to study the F-pattern of the stop consonants it is essential to take into consideration the F_2 peculiarities. The spectrographic analysis enables us to state the F_2 meanings with a vowel and their variational range for the stops. The meaning of consonant F_2 with a vowel defines the F_2 of the consonant itself with fair approximation, as the vowel a causes the least changes in the F-pattern of the consonant.

It must be pointed out that among localization series the defining of F_2 for the coronal (ʒ, c, ç) consonants proved of extreme complexity. Their meanings are stated mainly by formant transition and by comparison to other series may, thus, lack precision.

The obvious pecularity of the stop consonants' F-pattern is the following: the coronal consonants that are characterized by high F_2 with comparatively small range of meanings make one group. The other group is comprised of the labial and uvular consonants with low F_2 and the large range of its meanings. The velar consonants are distinguished by middle F_2 and the largest range of its meanings.

The ranges of the stop consonants' F_2 for male announcers produce the following meanings: (the figures in brackets show the F_2 consonants release in the front position of a vowel)

b-p-p̣	900 - 1900 Hz (1000 Hz)
d-t-ṭ	1400 - 2000 Hz (1650 Hz)
z-c-c̣	1400 - 2000 Hz (1600 Hz)
ʒ-č-č̣	1400 - 2000 Hz (1750 Hz)
g-k-ḳ	900 - 2300 Hz (1400 Hz)
q-q̣	750 - 1700 Hz (1100 Hz)

As is well known, the size of the consonant F_2 is determined by the length of resonator, the position of vocal chords, the degree of labialization and the degree of the rise of the middle of the tongue towards the hard palate (2). The F_2 of Svanian glottalized stops differs a little from the F_2 of the voiced and aspirated sounds. E.g., the F_2 of the glottalized consonants in the dental series is 100 Hz lower and, vice versa, is 100 Hz higher in the velar series. It is worth mentioning that the dental glottalized stop consonant is not affected by palatalization. E.g., the F_2 of ṭ sound in ä vowel proximity is 1600 Hz and that of d is 1950 Hz.

N. Chomsky and M. Halle (1) have observed that ejectives and palatalized sounds do not differ in the transition of second formant in the adjacent vowel. According to the authors, this peculiarity is caused by the shift of high-positioned glottis into neutral position. (It is commonplace knowledge that the glottis rising is necessary to create high intraoral supraglottal pressure). The fact that the similar features of second formant are not characteristic of Svanian glottalized and corresponding aspirated and voiced stops proves that the peculiarities caused by the tongue tension are more important in Svanian speech than those caused by the position change of the glottis.

References

1. CHOMSKY, N.--HALLE, M.: The Sound Pattern of English. New York, Evanston, and London, 1968.
2. FANT, G.: Acoustic Theory of Speech Production. Mouton, The Hague, 1960.

SPECTRAL FEATURES OF RENOWNED TENORS IN CD RECORDINGS

Geoff LINDSEY and David M HOWARD
Department of Phonetics and Linguistics, University College London, UK

Introduction

Many phonetic studies of singing make use of recordings made under laboratory conditions. However, vocal productions of the highest quality are found only under performance conditions. An early study [1] analysed such productions by using commercial recordings of a wide range of famous male opera singers. The present study takes advantage of modern digital technology, which has revolutionised both music recordings and speech laboratory analysis techniques.

Data Gathering and Analysis

It is essential, in the acoustic study of operatic singing, to distinguish the voice under analysis from any instrumental accompaniment. The analysis in [1] was carried out on unaccompanied notes taken from the recitatives and cadenzas of recorded arias. For compositional stylistic reasons such notes are often soft (quiet), short, and/or below the upper end of the singer's pitch range. The aim of the present work was to analyse several vowel qualities sustained loudly at high fundamental frequencies by celebrated tenors, and it was therefore decided to examine not an aria, but rather the cries of Vittoria! ('Victory!') by the character Mario Cavaradossi in the second act of Puccini's Tosca, many complete recordings of which have been made since the early fifties.

This vocal line, shown in Fig. 1, occurs at a moment of great dramatic intensity and is a well-known opportunity for vocal display. The score instructs the performer to sing fortissimo (very loudly), con grande entusiasmo, 'enthusiastically' and indicates that the orchestra should remain silent until the singer has ended the phrase. The word is sung twice; in its second, higher pitched, occurrence the first syllable is on F# (370Hz), and the rest of the word on A# (466Hz). It is this A# portion of the word, -toria, which was selected for analysis. In our data the two Italian singers (Giuseppe di Stefano and Luciano Pavarotti) perform this, as set by Puccini, with two syllables (the i only a glide), while the other singers clearly extend the i, providing a contrasting vowel quality for analysis. Although the -toria occupies less than a bar (and, at the tempo of the last previous metronome marking, would last 3.62 seconds), all the singers in our data sustain it for at least five seconds, one of them for well over eight!

The data were acquired with pre-emphasis onto a Kay Elemetrics DSP Sona-Graph (Model 5500) from commercial compact disc recordings, played on an Aiwa DX-660 compact disc player. The analysed performances were the following (all are analogue recordings which have been digitally mastered):

Giuseppe di Stefano: EMI CDS 7 47175 8 (1953)
Jussi Bjoerling: RCA GD84514(2) (1957)
Placido Domingo: RCA RD80105(2) (1973)
Luciano Pavarotti: Decca 414 036-2 (1979)
José Carreras: Deutsche Grammophon 413 815-2 (1980)

Narrow band (29Hz) spectrograms and long-term (>1s) power spectra were generated using the Sona-Graph Gray-Scale Printer (Model 5510). Fig. 2 shows the narrow band spectrogram of -toria sung by José Carreras in the most recent recording analysed, and Figs. 3 and 4 show the narrow band power spectra of the selected vowels: i for José Carreras and Jussi Bjoerling and o for all five tenors.

Discussion

Most noticeable in Fig. 2 is the regular vibrato, or fluctuation in fundamental frequency about the written note. The frequency of vibrato is between 4Hz and 5Hz for all the singers, and the width of the fluctuation range at the A# studied is between 40Hz and 50Hz for all the singers (200-250Hz at the fifth harmonic). Fig. 2 also clearly shows the distinct and prolonged medial vowel quality for i. The two surrounding qualities, for o and a, are much less distinct from each other. (We do not consider it appropriate to give phonetic transcriptions of the sung vowels: the perceptual effects of trained voice production on vowel quality are not under consideration here.)

The chief finding of [1] regarding the differentiation of opera singers from untrained singers was the high amplitude of a third formant between 2500Hz and 3000Hz (and in some cases the presence of a fourth) in the singing of the former. This high amplitude third formant probably corresponds to what has in the literature been called the 'singer's formant' or 'singing formant' [3]. Our data generally support this: the fifth harmonic (ca. 2330Hz) and/or sixth harmonic (ca. 2796Hz) are of high amplitude throughout. For low back vowels, the earlier study found this formant to be of lesser amplitude than the lower formants except in one of eight recordings; likewise, in the present data for o only Placido Domingo exhibits a fifth harmonic greater in amplitude than the other harmonics.

For all other singers the harmonic of greatest amplitude in o is the third (ca. 1398Hz) and even for Domingo this almost equals the fifth. Note that this is higher than the area of first and second formant energy that might be expected on the basis of average male values for [ɔ] or [ɑ] in speech as measured in [2]. At this fundamental frequency (ca. 466Hz) these would predict first and/or second harmonic amplitudes exceeding the third harmonic's. (In speech, an F2 around 1400Hz would be more appropriate for a centralised vowel quality.)

The greatest variation amongst the singers is shown above the fifth harmonic. Di Stefano and Domingo display substantial sixth harmonics (almost as high as the third and fifth in Domingo's case), but almost no energy above this. Both, but especially di Stefano, auditorily exhibit less "brilliance" or "shrillness" on this note than the other three singers. Pavarotti has a clear peak at the seventh harmonic, approaching the fifth harmonic in amplitude. (The width of the individual harmonics also reflects the width of each singer's vibrato: narrowest for di Stefano, widest for Pavarotti.)

Comparison of Fig. 4 with Fig. 3 shows that, in moving from o to i, both Bjoerling and Carreras have greatly increased the amplitude of the fourth harmonic (and to some extent the fifth), while decreasing that of the third (greatly in the case of Carreras, who also decreases the second). The change corresponds to the expected rise in F2 (c.f. [2]), and the greater acoustic change by Carreras is auditorily noticeable.

Acknowledgments

We are extremely grateful to B.M.G. Records (U.K.) Limited for making available the RCA recordings of Tosca; to Polygram Classics for the Deutsche Grammophon recording; to Allen Hirson and the Department of Clinical Communication Studies, City University, London, for lending us their Kay Sona-Graph; and to Hugh Webb for the use of his Aiwa CD player.

References

1. McGINNIS, C. S., ELNICK, M. and KRAICHMAN, M. : A study of the vowel formants of well-known male operatic singers. J. Acoust. Soc. Am. 23. 1951, 440--446.

2. PETERSON, G. E. and BARNEY, H. L. : Control methods used in a study of the vowels. J. Acoust. Soc. Am. 24. 1952, 175--184.

3. SUNDBERG, J. : The Science of the Singing Voice. DeKalb: Northern Illinois University Press. 1987.

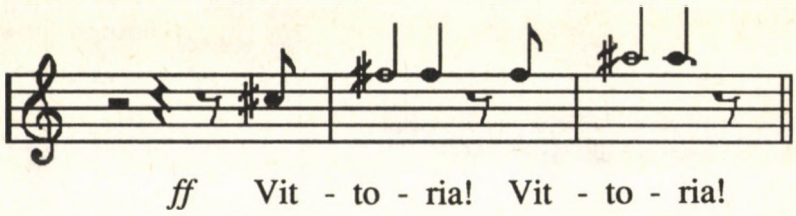


Fig. 1 The analysed vocal line, from Puccini's Tosca, Act II.

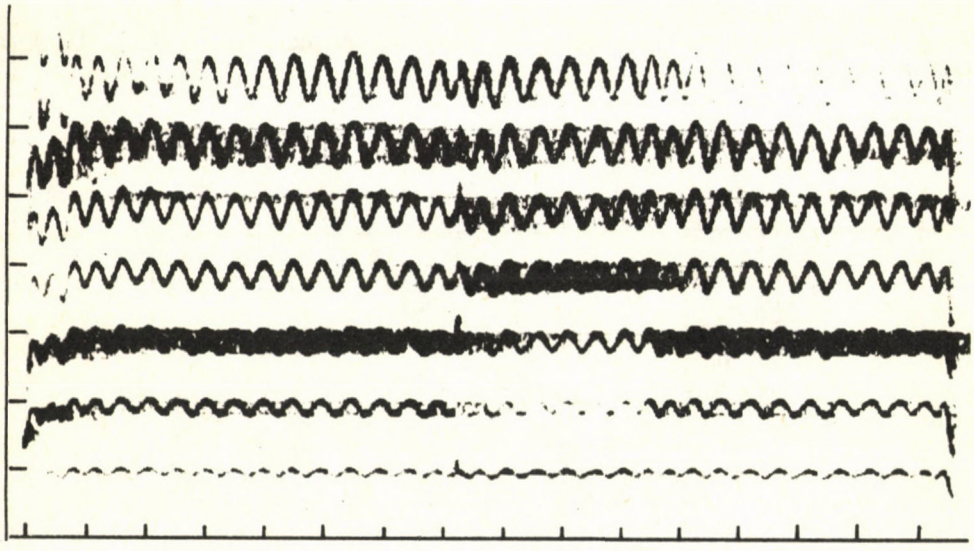


Fig. 2 Narrow band (29Hz) spectrogram of -toria (A#) in Vittoria: José Carreras
(Frequency axis: 500Hz/division; time axis: 400ms/division)

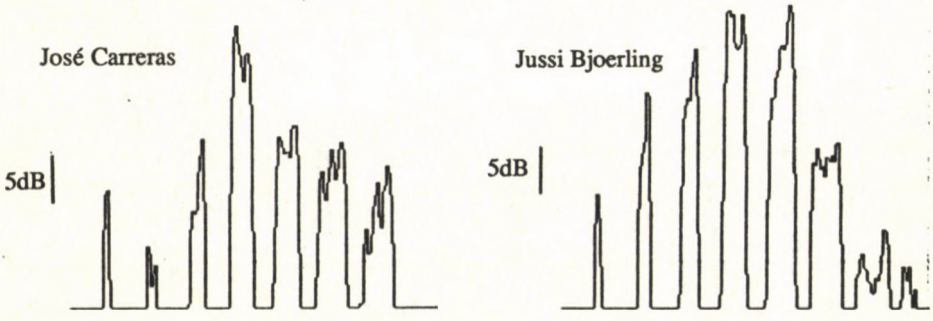
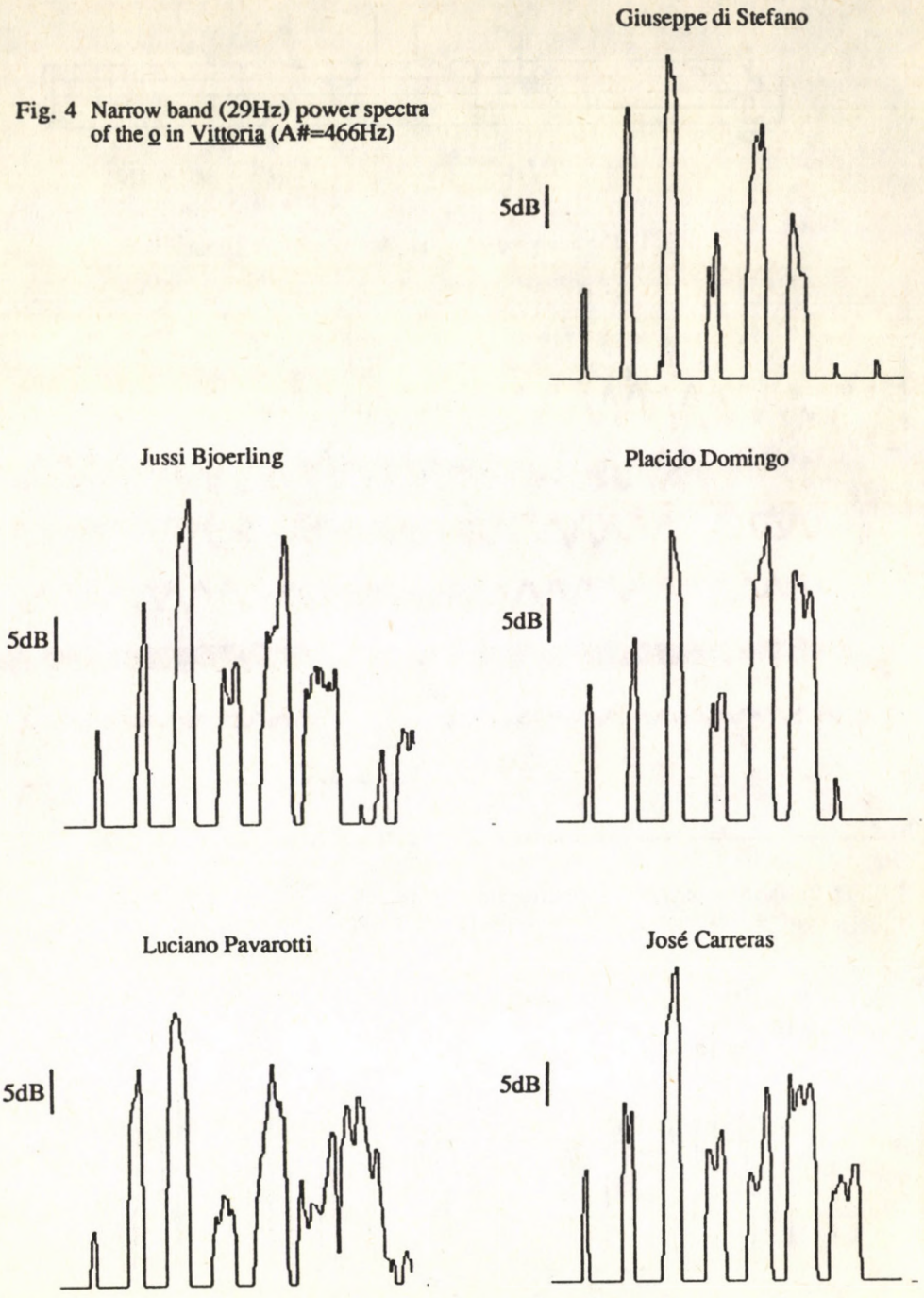


Fig. 3 Narrow band (29Hz) power spectra of the second i in Vittoria (A#=466Hz)

Fig. 4 Narrow band (29Hz) power spectra of the o in Vittoria (A#=466Hz)



COMPARISON OF SPECTRUM WINDOWS FOR SPEECH SPECTRA ESTIMATION

Henning REETZ
Max-Planck-Institut für Psycholinguistik
Nijmegen, NL.

Abstract

The reliability of different windowing functions in autocorrelation LPC analysis is compared. Synthetic and natural speech signals were analyzed sample by sample and the mean error and the standard deviations were computed. In sum, the Riesz window type performed slightly better than all other window types.

Introduction

Spectrographic measurements using autocorrelation LPC are usually done by placing a window on a part of the speech signal, computing the LPC spectrum and extracting the spectral peaks out of it as 'formants'. If the speech signal analyzed is reasonably constant, a relatively large window size is chosen (20 ms and more). In areas of rapid changes in the speech signal, smaller windows (down to 10 ms) are used. With a sampling frequency of 10 kHz, 10 or 14 poles are commonly found. Usually the window function is Hamming or Kaiser-Bessel. Those measurements are based on the assumption that neither placing the window to a slightly different position, nor choosing another window type has a critical impact on the computed data.

A very detailed and comprehensive discussion of windowing functions was done by Harris (1). We follow (1) with respect to all the terms and formulas throughout this paper. The outcome of this article is, that Kaiser-Bessel and Blackman-Harris windows perform best. This result has found its way into actual speech processing and especially the Kaiser-Bessel window seems to be replacing the Hamming window. It should be mentioned here, that (1) relied on concise theoretical considerations and an experiment involving the separation of two near-by sine tones.

Purpose

We have wanted to evaluate the influence of different window types in autocorrelation LPC analysis of speech signals. We generated several perfectly constant speech signals and carried out LPC analysis with several windowing functions moving them sample by sample over the speech waveform. The standard deviations (SDs) of the ascertained formant frequencies and bandwiths were computed and ranked.

Method

In sum, 101 speech signals were analyzed. In all cases the sampling frequency was 10 kHz. All signals were generated by copying one pitch period again and again, until the signal length exceeded 500 ms. Thus, all signals were perfectly constant. The 'mother' periods of 81 signals were generated with the Klatt formant synthesizer. Nine different formant patterns with nine different pitch frequencies were generated using the five formant cascade version with impulsive pulse source. Formant patterns and pitch frequencies were chosen following Monsen & Engebretson (2). The values are listed in table 1. The

peak amplitudes of the signals varied from 1800 to 2047 sample units. 20 naturally spoken vowels were added to this set. Two male and two female speakers pronounced the vowels [a:], [e:], [i:], [o:] and [u:] in isolation. These vowels were low pass filtered at 5 (five) kHz and digitized with a sampling rate of 100 (one hundred) kHz. One glottal period was extracted for each vowel and resampled with 10 kHz, such that its pitch period became a multiple of the 10 kHz sampling rate. These pitch periods were copied repeatedly until they exceeded 500 ms. The peak amplitudes of the signals ranged from 466 to 1977 sample units.

All 101 signals were pre-emphasized with a first order filter. The formant frequency estimation was done by autocorrelation LPC analysis and Bairstow root solving. Only the first three formant frequencies and bandwidths were evaluated. The data were analyzed with two different window sizes (10 ms and 20 ms), two different numbers of poles (10 and 14), and 8 different window types (Rectangle, Triangle, Riesz, Hanning, Hamming, exact Blackman, and Kaiser-Bessel with α=2.0, and α=3.5). For the synthetic speech signals the root solving was fed with the frequencies and bandwidths of the synthesizer, and for the natural speech signals with the output of a parabolic interpolation. We will refer to these frequencies in further discussion as 'intended' frequencies. The windows were moved sample by sample over the signals for one pitch period. For each individual window the formant frequencies and bandwidths were computed. Whenever the formant estimation failed, the values were ignored.

For each analyzing condition (no. of poles, window width, window type) and each signal the arithmetic mean and the SD of the frequencies and bandwidths were computed. The estimations were ranked. The SDs lower than 10 Hz were counted separately for each analyzing condition.

All computations were done with 32 bit integer, resp. 64 bit floating point (55 bit mantissa) arithmetic on a VAX in FORTRAN V4.7 under VMS. All programs were carefully tested. The window functions were implemeted following the formulas in (1).

Results

We did not get consistent results within the individual signals and analyzing conditions. For many signals, all windows but Rectangle had low SDs. Sometimes most windows had SDs close to zero, while some had SDs of several hundred Hz. Sometimes one or two windows had a SD near to zero, while the others had SDs of several hundred Hz. Because there's no room to present all results here, we will only state the main effects.

- For the synthetic signals, the differences between the intended formants and the computed ones were clearly influenced by the pitch and formant patterns of the synthesized signals and the number of poles in the analysis. In general, small SDs went along with small differences.
- For the natural signals, the computed values of different window types for one speech signal were nearly identical when their SDs were small.
- The SDs of the formant frequencies were lower than those of the bandwiths.
- The 20 ms windows gave less variant results than the 10 ms windows.
- The 10 pole analysis gave less variation than the 14 pole analysis.
- Higher pitch frequencies resulted in lower SDs.
- Frequency estimations of F3 were better than those of F2, which again were better than those of F1. The natural signals with 10 poles analysis did not fit into this pattern.
- For each window type the percentages of all SDs less than or equal to 10 Hz are listed below (c.f. Table 2. and Table 3.):

Synthetic signals:		Natural signals:	
Riesz	(79%)	Riesz	(58%)
Hanning	(75%)	Hamming	(44%)
Kaiser, α=2.0	(71%)	Hanning	(43%)
Hamming	(69%)	Kaiser, α=2.0	(38%)
Blackman	(62%)	Triangle	(32%)
Triangle	(57%)	Blackman	(30%)
Kaiser, α=3.5	(54%)	Kaiser, α=3.5	(24%)
Rectangle	(16%)	Rectangle	(16%)

Discussion

Placing a window at two arbitrary positions in a constant speech signal may lead to a difference up to several hundred Hz between the computed formant frequencies. As expected, 10 pole autocorrelation LPC performed better than 14 pole LPC for the 5 formant synthetic signals. Similar results were found for the natural signals. But we cannot conclude that 10 poles analysis should be preferred for natural speech signals, because we have investigated only a very small set of signals (especially no nasal or nasalized sounds). For the perfectly constant signals, larger windows improved the results substantially. Nevertheless, for some speech signals (both natural and synthetic) SDs larger than 1 kHz were found for all window types (20 ms, 14 poles).

As larger windows performed better, higher pitch frequencies led to better results, and higher formants were more reliably estimated, it is obvious that the amount of information within a window should be as large as possible. This is also supported by the fact, that the Kaiser-Bessel window with α=2.0 performed better than with α=3.5, since the latter is more 'slender' than the first. Taking into account that the rectangle window performed worst, we are back to the problem of finding a smoothly rising and falling window that does not suppress too much information.

In order to achieve reliable formant estimations we suggest that several spectrum estimations should be computed by moving a window over the part of signal of interest. Averaging the computed formants stabilizes the results. Moreover the SD of the formants is an indication of the reliability of the values found.

Finally, we want to stress the point, that results taken from 'simple' signals (e.g. sine tones) cannot be applied to the complex speech signal without critical discussion.

Further steps

We are investigating all window functions given in (1). We use natural speech signals, manipulated in the described manner, and speech signals that have not been manipulated. We are trying to find statistical measures, that can describe the differences between the variations better than the ones used here. We will try to get some insight into dependencies between characteristics of the speech signals and the variation of computed frequencies.

Table 1: Pitch frequencies and formant patterns of synthetic signals

Pitch frequencies [Hz]	[samples]	F1 [Hz]	B1 [Hz]	F2 [Hz]	B2 [Hz]	F3 [Hz]	B3 [Hz]	F4 [Hz]	B4 [Hz]	F5 [Hz]	B5 [Hz]
100	100	660	50	1720	50	2410	50	3900	300	4500	300
125	80	660	100	1720	100	2410	100	3900	300	4500	300
149.3	67	660	200	1720	200	2410	200	3900	300	4500	300
200	50	660	400	1720	400	2410	400	3900	300	4500	300
250	40	570	100	840	100	2410	100	3900	300	4500	300
303.0	33	490	100	1350	100	1690	100	3900	300	4500	300
344.8	29	270	100	2290	100	3010	100	3900	300	4500	300
400	25	300	100	870	100	2240	100	3900	300	4500	300
455.5	22	425	100	1230	100	2713	100	3900	300	4500	300

Table 2: SDs smaller than 10 Hertz, synthesized signals, in % (** stands for 100 %)

	10ms, 10 poles						10ms, 14 poles						20ms, 10 poles						20ms, 14 poles					
	F1	F2	F3	B1	B2	B3	F1	F2	F3	B1	B2	B3	F1	F2	F3	B1	B2	B3	F1	F2	F3	B1	B2	B3
Rectangle:	1	10	23	0	4	5	0	16	43	0	5	14	14	30	46	4	10	19	2	33	64	0	14	27
Triangle:	34	69	91	17	25	33	22	48	80	10	22	32	99	98	**	56	74	80	32	75	**	30	54	83
Riesz:	80	91	95	62	63	67	37	69	90	28	46	60	**	**	**	**	**	**	65	90	99	59	86	99
Hanning:	60	75	88	53	53	54	44	60	77	41	49	58	**	**	**	86	89	90	78	95	**	75	86	90
Hamming:	56	77	89	49	52	57	28	57	80	22	41	53	**	**	**	85	90	91	42	81	99	40	74	86
Blackman:	46	62	75	36	37	41	20	41	62	19	35	44	90	**	**	69	74	74	67	85	98	60	68	75
Kaiser, α=2.0:	49	74	86	48	51	51	38	54	77	38	43	49	99	**	**	83	88	88	74	88	**	68	78	86
Kaiser, α=3.5:	37	44	69	28	21	35	15	27	53	15	20	32	80	96	98	68	67	69	65	83	98	59	58	68

Table 3: SDs smaller than 10 Hertz, natural signals, in % (** stands for 100 %)

	10ms, 10 poles						10ms, 14 poles						20ms, 10 poles						20ms, 14 poles					
	F1	F2	F3	B1	B2	B3	F1	F2	F3	B1	B2	B3	F1	F2	F3	B1	B2	B3	F1	F2	F3	B1	B2	B3
Rectangle:	30	5	5	0	0	0	0	0	0	0	0	0	45	25	10	5	0	0	0	10	20	0	5	0
Triangle:	25	20	10	25	10	5	0	15	40	0	5	15	90	80	50	65	55	20	20	50	70	10	30	60
Riesz:	60	45	45	55	35	15	5	35	50	5	25	30	**	95	85	95	85	90	55	70	90	50	75	85
Hanning:	40	30	15	30	20	0	10	20	30	5	15	15	90	85	75	80	70	50	55	60	85	40	55	65
Hamming:	45	35	25	40	25	10	10	20	35	10	15	20	90	90	75	75	70	40	35	65	75	30	50	60
Blackman:	20	10	5	5	10	0	0	5	15	0	0	10	75	65	65	70	45	45	40	50	60	35	40	40
Kaiser, α=2.0:	30	25	10	20	10	0	5	15	20	5	10	10	90	75	70	75	55	50	45	55	80	40	50	60
Kaiser, α=3.5:	10	0	0	0	5	0	0	5	5	0	0	0	70	55	55	55	40	45	30	35	55	25	35	40

References

1. HARRIS, F.J.: On the Use of Windows for Harmonic analysis with the Discrete Fourier Transform. Proceedings of the IEEE, 66, 1978, p. 51-83
2. MONSEN, R.B. and ENGEBRETSON, A.M.: The Accuracy of Formant Frequency Measurements: A Comparison of Spectrographic analysis and Linear Prediction. Journal of Speech and Hearing Research, 26, 1983, p. 89-97

A METHOD FOR ANALYSING SINGLE VOWEL PERIODS

Tamás TARNÓCZY
Scientific Research Laboratories Hungarian
Academy of Sciences, Budapest, Hungary

Introduction

When fifty years ago I started to deal with Fourier Analysis by graphical method, the calculation up to 36 harmonies lasted more than 3 hours. In elaborating of my first publication (1941) I worked out 200 such analyses. The well known formulae of Fourier Pairs in complex form may be simplified when we use a summation instead of integral while the original time function F(t) will be approximated by n samples and converted into a step function $F(t_\mu)$. The new Fourier Pairs of

$$\left.\begin{aligned} F(t_\mu) &= \sum_{k=1}^{n-1} G(f_k)\, e^{j\omega k\mu/n} \\ G(f_k) &= \frac{1}{n} \cdot \sum_{\mu=1}^{n-1} F(t_\mu)\, e^{-j\omega k\mu/n} \end{aligned}\right\}$$

have discrete components both in the time and the frequency domain. Converting the formulae into real part alone and using trigonometric tables, it is possible to compute sine and cosine components up to n/2-th harmonic, i. e. half of the sampling rate (Nyquist limit).

The methods of frequency analysis became in the mean time easier by way of developing analog and digital band-pass filters. When an analysis is performed, there is a steady flow of time data into the filter which, in turn, produces a steady flow of filtered time data at the output, and the result is the convolution of the input signal with the impulse response function of the filter. This correspondens to a complex multiplication in the frequency domain of the spectrum of the signal and the frequency response function of the filter.

The technical process takes place in such a way that the time function (the oscillogram) moves many times through the filters of various center frequencies or, alternatively, only once through parallel filters. This procedure is too slow for many scientific purposes.

This situation changed drastically in 1966 when the time signal was first quantized (or sampled) according to a method named Discret Fourier Transform (DFT). At the same time the transformation algorithm became very fast through a new digital analysing procedure called Fast Fourier Transform (FFT). These two innovations together resulted in a quasi real-time analysis of both periodic and random acoustical signals.

Using this method the time function is time limited, sampled and windowed by a weighting function the so called "time window". The multiplication occure now in the time domain between the samples of the original signal and the appropriate

weighting function, The FFT itself is an algorithm working on digitized and stored data and using a limited number of multiplications and summations. The whole procedure looks like the old graphic analysis but with advantages of computer applications. The sampling function in time consists of a series of impulses. The multiplication of the two time signals actually corresponds -- according to the convolution theorem -- to the convolution of their respective frequency spectra.

The two novel ideas of this computation are that, first, the data of one group of signals could be processed as digital numbers of a common stored block and, second, that the number of computation steps will be reduced by a considerable degree. If we have n samples in the frequency domain, the normal computation algorithm involves n^2 steps. By a suitable arithmetical technique the steps could be reduced to n.ld n. This means at n = ld 2^{10} = 1024, an about 100 times shorter working time.

In a typical modern dual-channel signal analysier a frequency resolution for an upper limit of 6.4 kHz is 8 Hz, and for that of 3.2 kHz is 4 Hz obtained. These data are sufficient for attaining continuous spectra of consonants and about 3-4 formant places of vowels.

Experiments

At first, I made use of this scheme to obtain spectra of some noise-like speech sounds by means of a DFT/FFT analyser. The apparatus is able to average many spectra of various utterances spoken by the same person. Altogether, 52 tokens were recorded e.g. in such a word: [lɑ̊ʃ:ɑ̊n].

The second test was the analysis of vowels by the same method. It would be clear that such type of illustrations does not fulfill our expectations regarding the line spectra of a DFT analysis because the shapes and lengths of successive periods of the vowel are different. It is well known that during the utterance the pitch and timbre of vowel countinuously change. As our attempt to solve the problems I also tried to record averages of sung vowels having various pitches. However, this method was also unsuccessful.

Thus, the following new method was applied. First we determined the precise length between two sharp peaks of the period to be measured. Second we selected the next zero crossings, and made a quasi-infinite series from this period with the help of a computer. The next step was the analysis of this acoustic signal with a DFT/FFT apparatus through a Hanning time window. The result indicates the fundamental frequency of the selected period and specifies a strictly discrete line spectrum that suggests 3 to 4 formant regions. After obtaining a hard copy of this analyser, we repeated the process over consecutive periods for the total duration of the vowel. Fig. 1 shows an example for such an analysis of vowel [o:]. The measuring devices were B-K FFT analyser Type 2034, graphic recorder Type 2313 and H-P cumputer Type 2036 and plotter Type 2225.

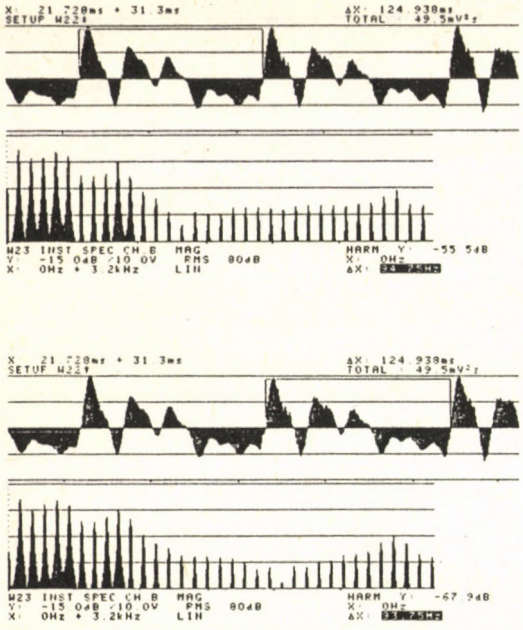


Figure 1. Oscillograms and spectra of two successive periods of a vowel [o:]

Results

I chose three two-syllabic words: [kåto:] a female name, [ko:tå] and [kot:å] possible utterances for music notes. In the diagram the abscissa represents the ordinal number of periods. The ordinate shows the frequencies of the fundamental (F_o) as well as of three formant centers. Talker was a male.

1. The character of the word intonation in affirmative form in Hungarian is always falling. This falling intonation can be followed numerically, e.g. the first [o:] in [ko:tå] falls from 183 Hz to 118 Hz (Fig. 2).

2. The stress always falls in the first syllable this vowel starts with the highest pitch, e.g. [o] in [kot:å] (Fig. 3).

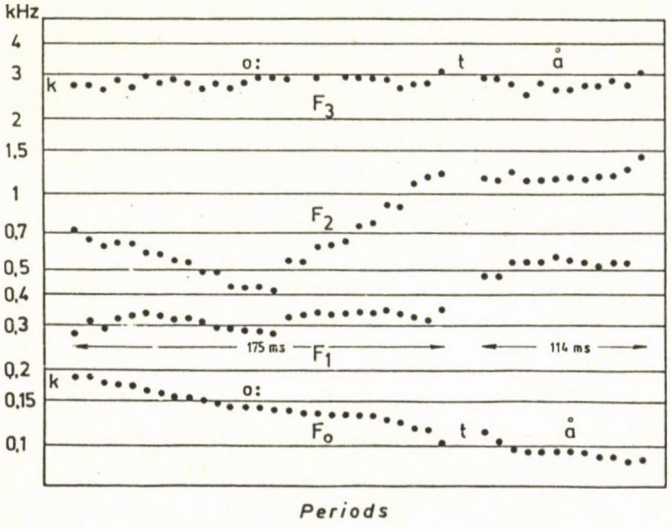


Figure 2. Results of the analysis of the word [ko:tå]

3. The second vowel follows the descending tendency of the pitch. In the word [kåto:] the first vowel falls from 117 Hz to 108 Hz, the second one follows it from 109 Hz to 87 Hz (Fig. 4). The largest change was measured on [ko:tå] the pitch of their vowels descends from 183 Hz to 84 Hz, exceeding one octave (Fig. 2).

4. Interesting results were also found in the formant patterns. In our analysis the motion of formants can be seen more precisely than in a sonogram. The transitions from one vowel quality to the next clearly show that no fixed formant places are characteristic to the various vowels. Only the third and the forth formants (later one not represented in the figures) remain stable -- a probable sign of the fact that these formants are talker-bound instead of deriving from vowel character.

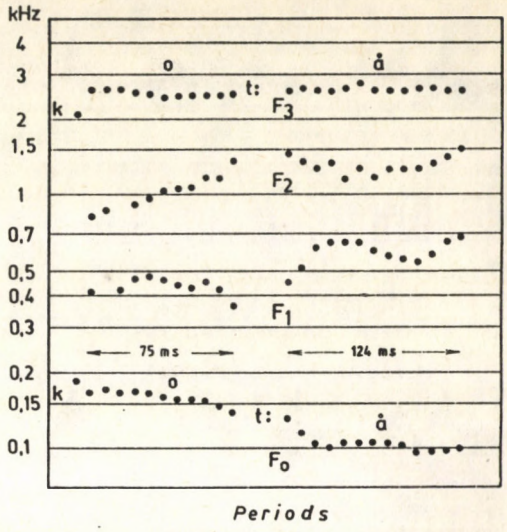


Figure 3. Reults of analysis of the word [kot:å].

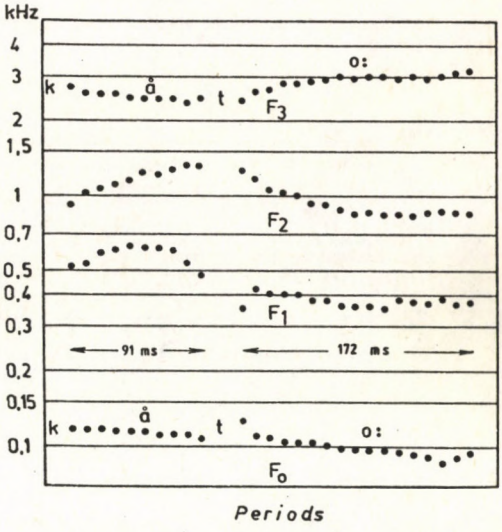


Figure 4. Results of analysis of the word [kåto:].

Acknowledgement

Further experiments with more talkers and more tokens are expected to bring further results. The author expresses his acknowledgement to Dipl. Phys. István Dániel for making the computer program and for his assistance during the experiments.

References

GADE, S. -- HERLUFSEN, H.: Use of Weighting Functions in DFT/FFT Analysis. Part I. Technical Review (B and K) 3/1987, 1-28; Part II. 4/1987 1-35.

ZUR AKUSTISCHEN STRUKTUR DEUTSCHER REIBELAUTE

Laszlo VALACZKAI
Phonetisches Labor der Attila-Jozsef-Universität Szeged, Ungarn

Die Studie behandelt die Bildungsdauer, die allgemeine akustische Struktur, die Angleichung der Frequenz- und der Intensitätsstruktur deutscher Reibelaute in weiblicher Aussprache.

Das Ziel ist die Ermittlung von Angaben für die interlinguale Analyse, vor allem aber für die Redesynthese.

Als Methode diente die Spektrographie mit den Instrumenten der Ungarischen Akademie der Wissenschaften.

[v] im Beispielwort "würdig":

1. Bildungsdauer: cca. 70 ms.

2. Allgemeine akustische Struktur: Die Intensität beginnt bei 0, nimmt während der ersten 30 ms 4 dB zu und erreicht nach weiteren 30 ms das Maximum bei 36 dB, das ist 14,5 dB weniger als die maximale Intensität des folgenden Vokals. F_1: 360 - 660 Hz, F_2: 1050 - 1720 Hz, F_3: 2370 - 3000 Hz.

3. Angleichung der Frequenzstruktur: Nur F_3 gleicht sich dem anschließenden Vokal geringfügig an: Wenn F_3 des Vokals unter 1500 Hz, etwa im Bereich von 950 - 1450 Hz liegt, erfolgt der Übergang zur Klarphase von cca. 1250 Hz.

4. Angleichung der Intensitätsstruktur: Es gibt keinen kontinuierlichen Übergang zur Intensität des folgenden Vokals, im Vergleich zum maximalen Intensitätswert des [v] tritt ein Rückgang von 4 - 5 dB ein.

[f] im Beispielwort "fad":

1. Bildungsdauer: cca. 110 ms.

2. Allgemeine akustische Struktur: Die Intensität beginnt bei 0, steigt unter Schwankungen von 2 - 4 dB bis 6 dB, nimmt cca. bis zur Hälfte der Bildung weitere 7 dB zu, erreicht einen Gipfel bei 13 dB, verringert sich während der nächsten 20 ms um 2 dB, erhöht sich während der folgenden 30 - 35 ms um 3 dB und steigt dann steil auf 19 dB. Das ist das Maximum, es liegt 27 dB niedriger als die maximale Intensität des anschließenden Vokals. Bis Abschluß der Bildung erfolgt ein Rückgang um 8 dB. Stimmlose Geräusche liegen zwischen 1300 und 8000 Hz.

3. Angleichung der Frequenzstruktur: Je höher F_2 des anschließenden Vokals liegt, desto höher liegt auch die untere Grenze der Geräuschkomponenten des [f]. Die höheren Frequenzen des Vokals passen sich den entsprechenden Bereichen des f an: wenn VF_3 über 4 kHz und VF_4 über 5 kHz liegt, kann der Übergang zur Klarphase des Vokals um 4 - 500 Hz niedriger einsetzen als der Gipfelwert des benachbarten Geräuschbündels.

4. Eine Angleichung der Intensitätsstruktur läßt sich nicht nachweisen. Die Intensität des Vokals zeichnet sich durch einen steilen Anstieg aus.

[z] im Beispielwort "Samen":

1. Bildungsdauer: cca. 80 ms.

2. Allgemeine akustische Struktur: Die Intensität beginnt bei 0, steigt während der nächsten 10 ms um 7 dB, fällt dann steil um 5 dB, erhöht sich allmählich innerhalb der folgenden 55 - 55 ms um 10 dB. Eine ausgeprägte Formantenstruktur fehlt, die Realisierung erfolgt im wesentlichen stimmlos. Die untere Grenze der Geräuschkomponenten liegt über 4 kHz.

3. Eine Angleichung der zwischen den Geräuschkomponenten und den Vokalformanten lassen sich nur zwischen 5 und 6 kHz nachweisen: die hohen Frequenzbereiche des V passen sich den entsprechenden Bereichen des [z] an.

4. Angleichung der Intensitätsstruktur: In der Anschlußphase erfolgt ein Übergang von cca. 16 ms, die Intensität nimmt 5 dB zu, von diesem Wert wächst die Amplitude in die des Vokals hinein, die sich rapid über 40 dB erhöht.

[s] im Beispielwort "Masse":

1. Bildungsdauer: cca. 150 - 155 ms.

2. Allgemeine akustische Struktur: Die Intensität beginnt in intervokaler Stellung bei 20 dB über 0 und schwankt mit Ausnahme der letzten 20 ms im Bereich von 18 - 22 dB. Während der letzten 20 ms tritt eine Verringerung um 5 - 6 dB ein. Intensive Geräusche liegen zwischen 1700 und 12000 Hz, der intensivste Bereich bildet sich zwischen 6400 und 7000 Hz heraus.

3. Angleichung der Frequenzstruktur: VF_2 beeinflußt die untere Grenze des Geräuschbündels: bei höheren VF_2-Werten liegt die untere Grenze der Geräusche höher und umgekehrt.

4. Eine Angleichung der Intensitätsstruktur tritt nicht ein. Die Intensität des anschließenden Vokals steigt steil an un erreicht den maximalen Wert innerhalb von 20 ms bei 50 dB.

[ʒ] im Beispielwort "Garage":

1. Bildungsdauer: cca. 70 ms.

2. Allgemeine akustische Struktur: Die Intensität des vorangehenden Vokals geht allmählich in die des [ʒ] über, dabei verringert sich die Intensität im Vergleich zum maximalen Wrt von 42 dB um 10 dB während einer Zeit von 50 - 55 ms. Die Realisierung des [ʒ] erfolgt unter einem starken Stimmtonverlust. Eindeutig läßt sich nur der Grundton von 230 Hz nachweisen. Zusammenhängende Frequenzbündel liegen zwischen 1,3 und 8 kHz.

3. Angleichung der Frequenzstruktur: Die Wirkung der benachbarten Vokale auf die Herausbildung der intensiven Frequenzbereiche des [ʒ] ist unvekennbar, diese Bereiche liegen zwischen 1300 - 2100, 2850 - 3450, 3700 - 4500 Hz. Übergänge der Vokalformanten zu den nächstliegenden intensiven Bereichen des Konsonanten erfolgen von 500, 1500, 2670 und 3800 Hz.

4. Ahgleichung der Intensitätststruktur: Der Übergang zeigt einen kontinuierlichen Amplitudenanschluß.

[ʃ] im Beispielwort "Schaf":

1. Bildungsdauer: cca.130 ms.

2. Allgemeine akustische Struktur: Die Intensität beginnt bei 0, steigt während der nächsten 20 ms um 2 dB, dann während der folgenden 50 ms um weitere 28 dB an. Der maximale Intensitätswert beträgt 32 db, das ist 10 - 12 dB weniger als die maximale Intensität des anschließenden Vokals. Von diesem Wert verringert sich die Intensität bis Abschluß der Lautbildung um 20 dB. Intensive Geräuschkomponenten erscheinen im Spektrum schon im Bereich von 1650 - 1850 Hz, zusammenhängende Bündel liegen jedoch zwischen 2370 und 7720 Hz.

3. Angleichung der Frequenzstruktur: Die labiale Koartikulation beeinflußt die Herausbildung der intensiven Stellen im Spekrum des Konsonanten, das gilt besonders für VF_2. Der Übergang zur Klarphase erfolgt cca. von 1550 Hz.

4. Eine Assimilation der Intensität tritt nicht ein.

[j] im Beispielwort "ja":

1. Bildungsdauer: cca. 100 ms.

2. Allgemeine akustische Struktur: Die Intensität beginnt bei 0 und steigt steil bis auf 35 dB. Nach einem ebenso steilen Rückgang von 3 - 4 dB tritt eine weitere Zunahme um 15 dB während der nächsten 40 ms ein. Der maximale Wert liegt 4 dB über dem Maximalwert des anschließenden Vokals von 48 dB. Formantenstruktur: F_1 200 - 560 Hz, F_2 2050 - 2550 Hz, F_3 2750 - 3200 Hz. (Etwas zu niedrig!)

3. Angleichung der Frequenzstruktur: KF_2 und F_3 werden nicht durch den Typ des nachstehenden Vokals beeinflußt. F_1 von [j] hat eine gewisse Ähnlichkeit zu F_1 des [I]; er steigt ebenmäßig zum F_1 des anschließenden Vokals.

4. Angleichung der Intensitätstruktur: Die Amplitude des j geht allmählich in die des Vokals über, die Intensitätszunahme beträgt dabei cca. 3 - 4 dB.

[ç] im Beispielwort "ich":

1. Bildungsdauer: cca. 220 ms.

2. Allgemeine akustische Struktur: Im Vergleich zum maximalen Intensitätswert des vorangehenden Vokals von 41 dB beginnt die des [ç] bei 21 dB. Während der folgenden 130 ms der Lautbildung tritt eine Verringerung um weitere 3 - 5 dB ein, dann ein Anstieg von 5 dB, und von diesem Wert verringert sie sich allmählich bis 0. Intensive Geräusche liegen zwischen 3 und 8 kHz. Die untere Grenze liegt bei 3 kHz, erhöht sich während der nächsten 65 - 70 ms bis 3,6 kHz und geht dann bis Abschluß der Bildung auf den Originalwert zurück.

3. Angleichung der Frequenzstruktur: Die Geräuschstruktur ist von der Formantenstruktur des benachbarten Vokals unabhängig.

4. Keine Angleichung in der Intensitätsstruktur.

[x] im Beispielwort "doch":

1. Bildungsdauer cca. 130 ms.

2. Allgemeine akustische Struktur: Im Vergleich zum maximalen Intensitätswert des Vokals von 40 dB beginnt die Intensität des [x] bei 9 dB, das ist 31 dB weniger. Intensive Geräusche liegen zwischen 3500 - 4300, 4750 - 5830 Hz, Geräusche von geringerer Intensität zwischen 4850 und 8000 Hz.

3. Angleichung der Frequenzstruktur: [x] gleicht sich dem Vokal nicht an; VF_4 paßt sich etwas der Struktur des [x] an.

4. Keine Angleichung der Intensitätsstruktur.

[h] im Beispielwort "Hammer"

1. Bildungsdauer: cca. 60 ms.

2. Allgemeine akustische Struktur: Die Intensität beginnt bei 0, erhöht sich während der nächsten 35 ms um 14 dB, dann steil um weitere 26 dB (das ist auch der maximale Intensitätswert des anschließenden Vokals). Die Frequenzstruktur wird stark durch die des folgenden Vokals beeinflußt.

3. Angleichung der Frequenzstruktur besonders durch VF_2.

4. Angleichung der Intensität: sie fällt im wesentlichen aus wegen einer Verringerung von 5 dB beim Übergang zum Vokal.

References

1. FANT, G.: Acoustic theory of speech production. - s' Gravenhage, 1970.

2. FÓNAGY, I. - SZENDE, T.: Zárhangok, réshangok, affrikaták hangszínképe, in: Nyelvtudományi Közlemények LXXI, 1969, 281 - 344.

3. NEPPERT, J. - PÉTURSSON, M.: Elemente einer akustischen Phonetik. Hamburg 1986.

4. OLASZY, G.: A magyar beszéd leggyakoribb hangsorépítő elemeinek szerkezete és szintézise, in: Nyelvtudományi Értekezések, Nr. 121, Budapest 1985.

5. OLASZY, G.: Die Anwendung des Flex-Deutsch Sprachsynthesesystems in phonetischen Forschungen, in: Magyar Fonetikai Füzetek 19, Budapest 1988 (Ed. by T. Szende), 34 - 46.

LISTENING FOR PHONEMES IN READING PROGRAMS

Anna ADAMIK-JÁSZÓ
Teacher Training College, Budapest, Hungary

In our college we wrote a book called The History of Hungarian Reading Instruction, in which my chapters concerned the period between 1868 and 1925 and the latest period after 1950. I intend to use these chapters in this paper.

Generally speaking, there were several changes during the last twenty years. We can prove that one change in itself can be useful -- and it was always proved --, but from the point of view of the whole system, every change caused damage, and finally the system was hurt. Let us enumerate the facts: a) first, the system after 1868 because we must know what was destroyed; b) then the marred system after 1950 and the destroyed one after 1969.

a) The Public Education Act in 1868 made public education compulsory and gave impetus to research and teaching. The curriculum in 1869 recommended the analysis-by-synthesis method, i.e., a combination which was developed in both German and Hungarian pedagogy. Its characteristics are the following: 1. Reading and writing were taught simultaneously; the written form of the letter was taught first, then the printed form. The pupils first read written words, later the printed ones. Actually they wrote first and read later. The purpose of this order was to teach blending. The official name of this method at that time was the reading-through-writing method. 2. The teaching process started with preparatory exercises, i.e., a foundation for both reading and writing. The preparatory exercises consisted of conversations on different topics organized according to the environment; of analysis of the speech into sentences, words, and sounds, and blending the sounds into words, sentences, and short stories. The analytic work prepared for the writing; the synthetic work prepared for the reading. The writing itself was prepared by drawing letter-elements. At the same time, the pupils were oriented in space learning, the directions. The preparatory exercises lasted six weeks, the real teaching of reading started after that. 3. Always the sound was taught first, then the corresponding letter. Every class was divided into two parts: sound-teaching and letter-teaching. Each had its logical steps. 4. These steps were the following: The teacher always started with a story, i.e., with the whole language. Discussing the story, he called the pupils' attention to a word containing the new sound which would be taught. Pupils pronounced the word, analysed it into sounds, took the sound in question out, articulated it loudly and correctly, and established its place in the order of the sounds of the word. After this process, the corresponding letter was taken. It was pronounced and blended with the already known letters. It was practiced on the chalk-board, on the printed tables hanging on the walls of the classroom, and finally in the primer. The words were read first by syllabication then in whole, altogether, always explaining the meaning. It is not true that this method did not take care of meaning. The word analysed had only one unknown sound/letter. This sample word at the very beginning was short, one or two syllables. The length of the sample words grew gradually. First the lower case letters were taught, each after the other, organized according to the difficulties of their shaping and pronunciation. Then the capital letters were taught.

A very special subject (embodying the Pestalozzian ideas), the speaking-thinking exercises, supported beginning reading. It developed the listening skills, corrected the defects of speech, enriched and corrected the vocabulary, developed abstract thinking with the teaching of semantic fields or webs. The speaking-thinking exercises were attached to other subjects abroad. Only in our country did they get a special position in the curriculum. They were taught in the first two grades, three classes per week.

Meanwhile, the method itself was corrected, i.e. alleviated. In 1899 a very Hungarian learning tool, the phonomimics was investigated which supported the teaching of the sound with signs made by the hand, and that of the letters with concrete pictures. The signs helped the blending, too. The phonomimics started with the printed letters, and this order remained until the present. The phonomimics worked well in practice. Unfortunately, it was discontinued in 1950 as a tool of "the bourgeois pedagogy".

b) In 1950 the phonic-analytic-synthetic method was introduced which did not differ from the method of 1869, only it lacked the playfulness of the methods utilized in the first part of the 20th century. Furthermore, the authorities allowed only one primer instead of the rich selection of literature of the previous period. Another big change was that the educational policy gradually diminished the number of classes of speaking-thinking exercises. Finally, they were cancelled in 1963. So, beginning reading instruction lost its important background. This step, undoubtedly, was the first one which damaged the system.

In 1969 the Budapest experiment was begun at the State Pedagogical Institute whose material was published in 1973 (A Fővárosi alsó tagozati kísérlet tapasztalatai, Bp., 1973. Szerk. dr. Hunyady Zoltán). Part of the experiment was the so-called functional method of teaching reading whose aim was to elevate the productivity without increasing the number of classes. Its characteristics are the following: The time of the preparatory exercises was reduced. After two classes, the letter-teaching started. The phonic, analytic, and collecting work of the preparatory exercises was considered too long and unnecessary. The authors introduced a psychological reason: first, the isolated letter-teaching hindered the understanding of the whole. Secondly, pupils wanted to read immediately; they wanted success at the very beginning.

The researchers shortened the time of the letter-teaching. They taught letter-groups instead of single letters. They taught every letter during the fall semester. They cancelled the sound-teaching then letter-teaching structure. The letter was shown, and after memorizing its shape, it was pronounced. In that way, the visual skills were emphasized instead of the listening, aural skills. Silent reading, an unknown technical term before in Hungary was emphasized. Syllabication was reduced at the same time. The active, individual work was emphasized using tests or worksheets. This was the period when workbooks were first used in reading. The result of the experiment is not the classical analysis-by-synthesis method, rather a synthetic one in which listening skills were pushed into the background. Writing was taught by imitation. The elements of the letters were not explained. This solution is understandable because the preparatory exercises were almost entirely cancelled. After the listening skills, the writing was hurt. We started to have the situation which was described by Jeanne Chall; the pupils colored, underlined, connected pictures, etc., on the worksheets instead of reading (Chall 1967).

The Ministry of Education introduced a new primer and a system of alternative programs after 1978. Alternative new primers were published

in 1980, 1985, and 1987. We must mention also that, during the 1970s, the global method was introduced for retarded children. The global method had an impact on the other programs of the 1980s. So we can establish that the four alternative programs were born under the influence of both experiment in 1969 and the global method. Let us see now briefly the characteristics of the four programs!

1. I Learn to Read (Olvasni tanulok) published in 1978 is a workbook whose authors are Mr and Mrs Romankovics and Dr Ildikó Meixner. The primer uses an analytic-synthetic method with global preparatory exercises. It means that, during the preparatory period, the pupils read whole words as in the look-and-say method. It was supposed that this global preparatory program provided an inspiring experience for the pupils. They taught only 39 whole words. The time of the letter-teaching was shortened. The lower case and capital printed letters and the lower case written letters are taught at the same time. The development of silent reading and worksheet exercises are central. Reading and writing are taught simultaneously. In the class-project, the first step (after the introductory exercises) is the letter--teaching which is followed by the recognition and analysis of the sound. The third step is the teaching of the cursive form of the letter. Finally, as the fourth step, they have exercises.

2. The Bear Is Reading (A maci olvas) by Dr Róbert Ligeti and Mrs Katalin Sahin-Tóth Kuti was published in 1980. This is the global program, but with phonics. It starts with 87 whole words; i.e., it starts later to teach letters and corresponding sounds. It does not teach blending at all. Its mastery is supposed to be achieved by the teaching of whole words. The teaching of writing starts later in the second semester. The authors quote the works of the Belgian Ovid Declory in which child perceive first the whole and do not analyze. Children are not able to do the early analysis. They emphasized playfulness, too.

3. An experiment, which lasted almost two decades, was supported by the state, and had a special program for the whole lower section, i.e., for the 1--4 grades of the elementary school was the Language-Literature-Communication program originated by József Zsolnai. It was officially introduced in 1985. It presents no special reading program for the first grade, but it uses both a synthetic one and a global one simultaneously. It means that they have a primer Éva and Feri Learn to Read (Éva és Feri olvasni tanul) which is a workbook for the global program, and they have another primer (Betűről betűre) for the synthetic program. They are bound into one volume titled Word and Letter (Szó és betű). How does this work in practice? The teaching starts with a global preparatory section, and the letter--teaching starts later, similar to the first program. The teacher follows a model of letter-teaching, whose steps are the following: a word is taken, then a letter is taken out. The letter is named by the teacher. He can pronounce the letter, but it is not compulsory. Speed reading training is used and the so-called eye movement training. It was proved that eye movement depends on comprehension; it can not be improved using drills independently (Pearson 662). They teach blending only for those who need it. They use silent reading - oral reading order. They read phrases for the sake of the later teaching of the grammar.

4. The fourth program is the intensive-combined program, whose author is Dr Gabriella Lovász. It does not use preparatory exercises at all, neither global nor classical ones. She starts to teach the letters immediately using letter-groups by a synthetic method. After the first overview, she teaches again the letters simultaneously with writing. She brings great care to bear upon sounding and forming the sound-letter correspondences using

pictures. Silent reading followed by oral reading order is used. Syllabication is not used at all, similar to the other programs. The most important grammatical terms are taught such as sentence, word, sound, vowel, and consonant. The title of the primer is Letter-Fair (Betűvásár). It is a workbook, too. Blending is taught very carefully, which is an advantage of this program, however, it is considered too fast at the beginning.

Conclusions. The authors of the different programs and other authorities have made evaluations. They have proved that the programs work well in practice. (A tanító, 1984, 1.sz. 4-9, 2. sz. 1-10, 3. sz. 1-11). Contrarily them several researchers report that difficulties have been arising, especially at the end of the first schoolyear and later in the advanced grades. The teachers at the university for education of defective children report the increasing number of dylexics or pseudo-dyslexics. In my opinion, all the four programs must accept the responsibility for this situation.

I can summarize the reasons in three points: 1. the lack of the preparatory exercises, 2. the cancellation of sound-teaching as an independent step in the process, and 3. the total lack of syllabication. I can continue this list with the dominance of silent reading and the use of worksheets where the pupils do everything except reading. The whole list is connected with the listening skills damaged. In that way the programs can have success at the beginning of instruction, but the whole collapses because the system of subskills has not been built up. Our situation is similar now to the situation of the history of American reading during the 1960s when the big examinations investigated the failure of the look-and-say method. It is not an accident that the system of subskills was emphasized during the 1970s, and now the whole language approach, simultaneous reading and writing, reading aloud, phonics, and above all the metacognitive skills are in the center of research (see i.e. Pearson 1984 pp. 353-395, Chall 1967, Chomsky 1970, Samuels 1976, Adams 1978; this last one is a manual for the teaching phonics and writing).

Now there is a saying among educators; beginning reading instruction is not so important because it can be corrected later. But we all know that it is impossible or very difficult to correct it. On the contrary, not only reading has troubles, but grammar and the development of whole language instruction, because reading is not only reading, but the foundation of the whole-language development and the important base for other studies.

References

1. Adams, A.H.: Success in Beginning Reading and Writing. Good Year Books, Glenview, Illinois, 1978.
2. Chall, J.S.: Learning to Read: The Great Debate. McGraw-Hill, New York, 1967.
3. Chomsky, C.: Reading, Writing, and Phonology. Harvard Educational Review 40. 1970, 287-299.
4. Pearson, P.D.(ed.): Handbook of Reading Research. Longman, New York & London, 1984.
5. Samuels, S.J.: Hierarchical Subskills in the Reading Acquisition Process. In: Aspects of Reading Acquisition. Ed. by J.T. Guthrie. The John Hopkins University Press, Baltimore&London, 1976.

IS A UNIVERSAL PHONETIC STANDARD POSSIBLE?

Kálmán BOLLA
Department of Phonetics, Loránd Eötvös
University, Budapest, Hungary

Introduction

Every discipline seeks to establish an objective and exact measuring system for the substantive description of its systematic structure-building elements. The minimal linguistically relevant units of a phonetic system, i.e. the speech sounds, are usually represented by conventional graphic symbols. Phonetic transcription raises a number of important problems in terms of both the shape and the interpretation of symbols. Such problems have prompted the author to deal with the exact measurability of the phonetic quality of speech sounds.

Methods

Phonetic features determining the quality of a speech sound can be investigated in three aspects: articulatory, acoustic, and perceptional. Each aspect has an appropriate set of instruments and methods of investigation associated with it. For determining the articulatory features of speech sounds, I used labiography, palato- and linguography, and cineradiography. The acoustic parameters were analysed by dynamic sound spectrography, oscillography, and sound synthesis. Perceptional information was gained by auditive testing of both segmented (natural) and artificial sounds.

Discussion

The determination of the phonetic quality of speech sounds even today takes place mainly on the basis of perceptual impressions, by way of comparing with sound patterns (stereotypes first of all the sound types of the mother tongue) fixed in the investigator's mind. It is often found that the same speech sounds are differently characterized by different authors, marked with different phonetic symbols, thus made seem different sound units. The opposite thing also happens; a graphic symbol can cover sounds of highly different qualities. The confusion is even greater if the authors do not make a distinction between phonetic and phonematic types of analysis. In interlingual comparative and typological phonetic investigation this leads to false conclusions. In my opinion it is very urgent to solve the following three problems:

a) The criteria of the linguistic-phonetic segmentation

and dissection into constituting units of the sound stream should be made clear in order to make objective distinctions even in the problematic cases of segmentation, to be able to distinguish a diphthong from a phonetic juncture, a long consonant from a geminate, transition from a linking vowel etc.

b) The quality features of speech sounds (both the articulatory and the acoustic features) have to be revised, completed and made more unambiguous. A universal system of quality features has to be elaborated which would make a net (a measuring grid) for the qualification of sound units of any language.

c) The exact methods of quality feature measurement have to be determined and made generally accepted by phoneticians. Only then can a more exact correlation of characteristic features of production, phonation and perception be determined; besides, thus we could make a step forward in the exploration of the system of connetions of phonetic quality features and phonological distinctive features; in laying the phonetic foundation of the theory of phonology.

On a more solid ground we would be able to investigate the processes of the phonetic and phonological coding of language more effectively.

For lack of space here I can only deal with the problem of the quality of vowels. The phonetic features relevant from the linguistic - phonetic point of view are proper to be given in the following systematical arrangement:

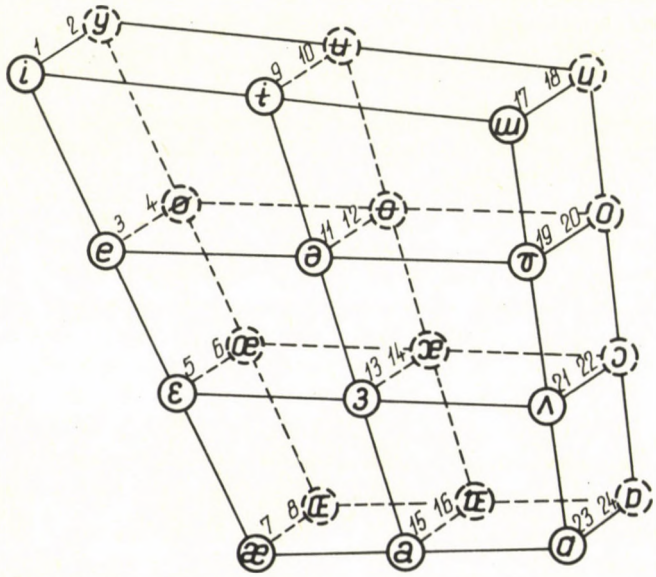
1. Vocalic--consonantal character
   1.1. pure vocalic 1.2. semi-vocal 1.3. consonant
2. Direction of sound-producing air stream
   2.1. egressive 2.2. ingressive
3. Force of articulatory activity
   3.1. tense (fortis) 3.2. normal 3.3. lax (lenis)
4. Homogeneity of articulatory process
   4.1. monophthong 4.2. diphthong 4.3. triphthong
5. Vibration of vocal cords
   5.1. voiced 5.2. media 5.3. unvoiced
6. Duration of articulation
   6.1. long 6.2. half-long 6.3. short 6.4. reduced (ultra-short)
7. The position of the soft palate
   7.1. oral 7.2. naso-oral 7.3. nasal
8. Horizontal movement of the tongue
   8.1. front (palatal) 8.2. central 8.3. back (velar)
9. Vertical movement of the tongue
   9.1. high 9.2. mid-high 9.3. mid 9.4. mid-low
   9.5. low
10. Opening of jaws
    10.1. close 10.2. half-close 10.3. half-open
    10.4. open
11. Articulation of the lips
    11.1 unrounded 11.2. rounded 11.3. protruded

It is a more difficult task to define these different articulatory features of sounds. It is still more difficult

to find the acoustic equivalents (projections) of features of articulation and to demonstrate their effect on perception. With knowledge of correlations the phonetic quality of a speech sound can be more firmly stated. There are some generally used methods for demonstrating and measuring sound formational features, nevertheless, there does not exist a universal standard of measurement, though there is a great need of it.

Each of the 11 different aspects suggested for the articulatory definition of sound quality demands a different measuring system. There are some features which can universally be measured in an exact way, by absolute parameters (e.g. voiced--unvoiced, oral--nasal, palatal--velar), while others can be interpreted by relative indexes, by their internal proportions (e.g. short--long), occasionally functional aspects must also be taken into consideration (e.g. semi-vowel--consonant), and in most cases a certain feature is realized dispersed in a broader or a narrower range. Sound quality is determined by the total of phonetic features, in a most peculiar mixture. We often find that sounds having the same main articulatory features greatly differ from one another. It is also well known that in different languages there can be fundamental differences between the physiological processes resulting in the nasality of a vowel.

I think that for the determination of sound quality more quality features must be taken into consideration than the amount generally used for the characterization of speech sounds.



Finally I present a method for the measuring of sound specifications enumerated as items 8, 9, 11, and partly 10. Its main point is a measuring grid which contains the above mentioned features of articulation and parameters F1 and F2 which can be correlated with them (see fig.). I have given the exact physiological characteristics, cineradiogram and measured data, the acoustic parameters for the place of articulation and the data of artificial production of the sounds that can be produced at the corner points of the articulatory chart (24 vowels). Thus the basis necessary for the comparison of speech sounds and for the representation of the stock of sounds and the sound-system is obtained for the three phonetic levels (articulation, phonation and perception). (In greater detail, see the publications listed in the References.)

Conclusion

There is a need for the measuring of the phonetic quality of speech sounds, the systematization of sounds on the basis of features of articulation and a more exact method applicable for interlingual phonetic comparisons. There is a demand for a universal standard which could be used for the measurement of sound quality and the systematization of sounds on the basis of their phonetic properties and which could be widely used. The preliminary conditions of its elaboration have been created.

References

BOLLA Kálmán: Egyetemes fonetikai hangszabvány? A magánhanghangzók. (A universal phonetic standard? Vowels). MFF 13. 1984, 71--120.

BOLLA, K.: On the Measurement of the Phonetic Quality of Vowels. Is a Universal Phonetic Standard Possible? AUB S-Philol, SLingu XV, 1984, 41--54.

The Principles of International Phonetic Association. London, 1949/1970.

COMPUTER PROGRAMS WITH SPEECH OUTPUT IN TEACHING READING AND WRITING - EXPERIENCES AND RESULTS

Iréne DAHL*, Karoly GALYAS
Dept. of Speech Communication & Music Acoustics, Royal Institute of Technology (KTH), Stockholm, Sweden
*also Dept. of Psychology, University of Umeå, Sweden

Theoretical background

As reading and writing are complex processes, it is impossible to designate an unambiguous cause of difficulties experienced in reading and writing (dyslexia). However, various forms of linguistic deficiency - of which the most apparent is an inability to segment words into phonemes - have frequently been found in school children suffering from difficulties in reading and writing (2)(3)(4)(5)(9)(10).

There is no one-to-one relationship between the graphemes (characters) of the alphabet and the phonemes (sounds), which they represent, as the sounds of a language adapt themselves to their linguistic surroundings, e.g., the "p" in the Swedish word "spik" (nail) sounds closer to a "b" than a "p".

The segmentation of words into their individual phonemes, their identification, and association to the correct grapheme and/or achievement of the coarticulation of individual phonemes, demands a shift in the attention from the content of a word to its form. This shift in attention demands a special cognitive ability. That this causes difficulties is revealed, e.g., by the fact that many children with difficulties in reading and writing experience difficulties in rhyming (8). The association between difficulties in reading and writing and low self-confidence is widely accepted and scientifically documented (6)(7). The OVE-project which started in 1983 is a cooperative, interdisciplinary project. The project is established at the Royal Institute of Technology (KTH) in Stockholm with Karoly Galyas as a project leader and responsible for the engineering aspects of the project, and at the Department of Psychology, University of Umeå, where Iréne Dahl is responsible for the methodology and pedagogical design of the programs. The hardware was based on the multi-lingual speech synthetic system developed at KTH (1).

Objectives and Hypothesis

Our objectives were to develop pedagogical programs (based on synthetic speech) to train the linguistic awareness and to study the effects of this new way of working for school children suffering from difficulties in reading and writing.

We assumed that simultaneous auditory and visual feedback will enhance the perception of the association between sounds and graphemes and that this enhanced perception will lead to an increased cognitive ability to absorb the form of the language.

We have developed programs for training sound discrimination, phoneme identification, letter-sound relationship, and spelling. A simple word processing program is used to write texts with phonemic feedback of the keystrokes. Words, sentences, or the entire text can be spoken.

Implementation

Up to and including the start of the spring term of 1987, we used a small lap-top computer. Its memory capacity was limited and the display could show only four lines, each of 20 characters. The programs were loaded and overlaid from a built-in cassette recorder. This computer featured a built-in printer. We also had access to a larger, external printer.

Nine children (all boys between 9 and 12 years) participated in the first, exploratory study, which was concluded in March 1984. During the spring term, they spent two 20-minutes sessions a week over a period of 12 weeks with the training program. Pre- and posttesting gave such promising results, that we were stimulated to continue with this work.

The small display offered many advantages when we worked with the training program but, as the abilities of the children improved, the demand grew for a larger screen. In the autumn term of 1987, the project grew to encompass five teachers at three schools and 12 children - five boys and one girl from the primary school and five boys and one girl from the secondary school. The programs had now been ported to an IBM-PC compatible computer with a large colour screen, and the duration of the children's sessions was increased to two 40-minutes sessions a week.

Software

The software comprises a writing program (TYPE AND LISTEN), which can be likened most closely to a speaking word processor and 15 different training programs, see Fig. 1.

<1> TYPE AND LISTEN

<2> SPELLING
<3> WHICH SOUND COMES FIRST?
<4> WHICH VOWEL IS IT?
<5> FIND ALL THE VOWELS?
<6> LONG OR SHORT VOWEL
<7> PAIRS OF LONG AND SHORT VOWEL
<8> LONG VOWEL BECOMES SHORT
<9> WHAT IS THE WORD (SYLLABLES)
<10> WHAT IS THE WORD (PHONEMES)?
<11> WHAT IS THE SOUND?
<12> WHAT IS MISSING?
<13> BUILDING WORDS WITH -ING
<14> BUILDING WORDS WITH -FUL
<15> RHYMING
<16> HANGMAN
<17> WORDLISTS
<18> QUIT

Fig. 1. The menu as it is presented on the screen.

Type and listen

After each keystroke, a spoken sound will be heard which corresponds to a phoneme, i.e., not to the name of a letter. When a word has been completed and the space bar pressed, the entire word will be spoken. The function keys are used to request the pronunciation of the current sound, the previous sound, the word, the previous word, the current sentence, the previous sentence, and parts or the whole passage. The desired function can be repeated as many times as the user wishes. The phoneme feedback can be omitted or changed to spelling.

The various programs train, for instance, the discrimination of homophones, the identification of the vowel or vowels in words, long, and short vowels, synthesis of words when pronounced syllable-by-syllable or sound-by-sound, positional analysis (identification of the location of a particular sound in a word), the subtraction of sounds (the removal of part of a word; the child must determine what has been removed), rhyming ability, and the conversion of the infinitive forms of verbs into nouns ending in -ing and -ful, e.g., write - writing. One of the programs comprise the old favourite "Hangman" in which the child, starting with the number of letters in the word, receives a certain number of guesses to fill in the word. This game is used to build up a knowledge of the "anatomy and structure" of words.

Instructions to the child and auditory feedback (When depressing alphabetic and numeric keys) are issued in the form of synthetic speech. Thus, the children are not required to read in order to understand their tasks or to use the programs. There are two printer programs. One prints the original passage and the other prints all keystrokes (i.e., logs the child's keystrokes). The original passage and the keystroke log can be stored on a discette.

<17>WORDLISTS allows the teacher to create new lists of training material for any of the exercises.

Method of working

The method of working ensures individual training for each child. The child types a passage. The edited passage is the child's property and receipt for the word known. The keystroke log file can be used for diagnosis and to assist the teacher in studying how a child has worked and to determine any areas in which he/she needs special training. The training programs include suggested word lists. The teacher can, however, add the words or sounds that he/she considers the child needs to practice.

The "mini word processor" does not include a spelling checker. The cursor-control keys enable the child to move around in the text passage, and the function keys permit the reading of individual words, sentences, parts of the passage, or everything that has been typed. The auditory feedback offers, for instance, an improved chance for the child to detect and correct his/her errors. The corrective part of teaching is, therefore, largely performed by the child. The teacher's role of forestalling errors and assisting the child to find his/her own approach to problems is accentuated.

The writing program can also be used to practice reading. All passages typed by the child can be stored on diskette. The teacher can type new passages that are unknown to the child. After having received background information on the content, the child can read the passage from the screen. If the child experiences difficulties in reading a word, the machine can be asked for the correct pronunciation. The child can request the reading of individual words, sentences, some long portion, or the entire passage.

Evaluation

Although no summary yet is available of the results of the pre- and posttests and of the interviews and questionnaires of the latest evaluation, the outcome appears promising. In parallel to the evaluation of the new version of the software, an investigation was also conducted with 100 children in the second grade (8-9 years old) to obtain an idea of the effects of simultaneous auditory feedback on their spelling skills.

For this investigation, three word lists (each of 20 words of equal difficulty) were dictated. One test comprised conventional dictation with pen and paper. The second list was typed on a computer keyboard without auditory feedback, and the third list was typed on a computer keyboard with simultaneous auditory feedback. For each dictation test, the child was alone with the teacher. Form teachers and special teachers completed an assessment of the normal performance of the class in class teaching.

Results

The results of the spelling-test investigation indicate the very positive effects of auditory feedback on the entire group and, especially, on the low-performance children, see Fig. 2.

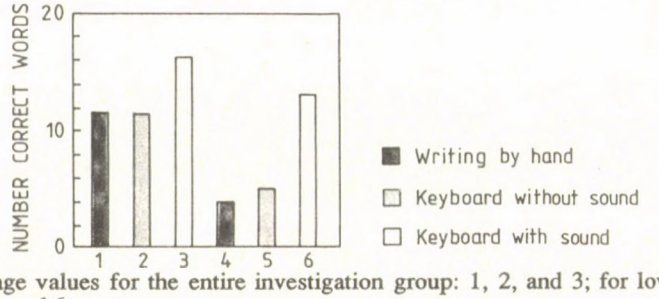


Fig. 2. Average values for the entire investigation group: 1, 2, and 3; for low-performance children: 4, 5, and 6.

The number of children scoring 6 or less in the normal dictation test was 12, and they are classed as low-performance children.

When taking dictation by hand or by computer without auditory feedback, the average values coincide for the entire group, whereas they differ for the low-performance children. None of the children had any previous experience of the keyboard used with the computer. The slightly improved results of the low-performance children when typing at the computer can indicate that they also suffer from difficulties of manual coordination. The use of auditory feedback increased the performance of both groups and for the low-performance children, the difference in the average scores (from 3.75 to 13.08) gave an improvement of more than 300%.

Experience

Of the considerable experience we have gained over the past few years, we have only space to review a small part here.

The method of working with the talking computer (which we call OVE) gives the children a new approach to coping with their difficulties. Auditory feedback of what is being typed offers unique ways of:

*experimenting with sounds
*identifying and consolidating sounds
*practicing analysis
*practicing synthesis
*practicing the distinctions between long and short vowels
*detecting omitted or transposed letters
*self-discovering and -correcting their errors

We have noticed that the above has marked effects on the children's motivation and the mobilization of their resources.

Time is utilized very effectively. There has also been a noticeable improvement in the self-esteem for the children that have had access to OVE for an extended period.

References

1. CARLSON, R., GRANSTRÖM, B. -- HUNNICUTT, S.: A multi-language text-to-speech module. 1604--1607. Proc. ICASSP 82 3. Paris, 1982.
2. LUNDBERG, I.: Språk och läsning. Malmö, 1984.
3. LUNDBERG, I.: Lack of phonological awareness. A Critical Factor in Developmental Symposium. Wenner-Gren Center, Stockholm, 1988.
4. OLOFSSON, Å.: Phonemic Awareness and Learning to Read: A Longitudinal and Quasi-Experimental Study. Thesis. University of Umeå, 1985;
5. SKJELFJORD, V.: Teaching children to segment spoken words as an aid in learning to read. Journal of Learning Disabilities 9. 1976, 297--306.
6. TAUBE, K.: Reading Acquisition and Self-Concept. Thesis. University of Umeå, 1988.
7. TAUBE, K.: Läsinlärning och självförtroende. Rabén & Sjögren, Kristianstad, 1988.
8. TORNEUS, M.: Rim eller reson. Direktrapport 2. PsykologiFörlaget AB, Stockholm, 1983.
9. WAGNER, K.R. -- TORGESEN, J.K.: The nature of phonological processing and its causal role in the acquisition of reading skills. Psychological Bulletin 101. 1987, 192--212.
10. VELLUTINO, F.R. Dyslexia. Scientific American 3. 1987, 20--27.

SPEECH INTELLIGIBILITY EXAMINATIONS IN LECTURE HALLS WITH DIFFERENT METHODS AND THEIR RESULTS

András KOTSCHY - György SIMON

ÁÉTV-Building Design Institute, Acoustic Lab.
1016 Budapest, Krisztina krt. 99. Hungary
Hungarian Dubbing and Video Studios
1021 Budapest, Vörös Hadsereg u. 64. Hungary

Introduction

Recently we have carried out room acoustic measurements of wide range in school-rooms, in university lecture halls as well as in the halls of new cultural complexes. The objective is to make preparatory proposals for standardization and designing tendencies for small and medium size halls of public buildings.
This research work was co-ordinated by the Faculty of Building Structures of Budapest Technical University, Hungary.

Methods

We have done speech intelligibility tests according to the Hungarian Standard MSZ-3392-54 in five lecture halls with 20 persons and on one occasion with 40 persons, seated in small groups at different points of the halls. Two of the standard Hungarian speech intelligibility text samples were used. The text samples contain CVC (consonant vowel consonant) nonsense words of 200 syllables (47x one, 42x two, 23x three syllables) where consonants and vowels appear with the same probability as in the normal language. One of the samples was read out by an amateur actor with good pronunciation and the other which had been previously recorded in an anechoic chamber was played in by a tape recorder. At the recording the text samples were read out column by column alternately by an amateur actress and actor with good pronunciation (in the text samples there are 10 columns).

At the examinations the sound pressure level of the uttered logatoms is a very critical parameter. To decide the optimal value, we have taken into consideration the average speech level of the lecturers. When playing in through loudspeakers, we have used a higher sound pressure level than in the case of natural (live) speech. In each case, we have measured the average sound pressure level at a distance of 2,5 m from the speaker or loudspeaker.

In our measurements results, the logatoms written down without any mistakes are given in percentage of all the uttered logatoms. This is the phonetic balanced (PB) word

score.

In the recent past new instruments have been developed by the Brüel & Kjaer for a new objective method to measure speech intelligibility, which gives rating in less than 10 seconds instead of the hard and long work of the traditional examination. This is the Rapid Speech Transmission Index (RASTI) method. With the kind help of the Brüel & Kjaer, we had the possibility to carry out the speech intelligibility measurements with their Speech Transmission Meter (Type 3361) in the five lecture halls that had been examined previously. It offered a good possibility to compare the results of the traditional methods with that of the new objective ones.

The results and their evaluation

In Table 1 the speech intelligibility data of the five examined halls are given with some further parameters such as seating capacity, volume and number of test persons.

Table 1

Halls	seating capacity	V (m^3)	number of test persons	speech intelligibility (%) natural speech	artifical play-in	RASTI method
BME E/A	221	1350	20	94,3	83,8	75,0
BME E/B	408	2280	20	92,9	80,8	72,0
BME KII/32	221	760	20	92,4	83,7	74,0
KMF I	156	820	40	91,0	85,2	70,0
BME KII/53	288	1460	20	89,8	82,9	74,0

The PB word scores of the examined halls are less different from one another. In our opinion, one of the reasons is that the text samples are not difficult enough. There are too many words of one syllable, the article "a" occurs frequently, etc. Here the mistakes can be due to inattention only.

The most frequent misspellings of the text samples are the following:

(i) recording phonetically cognate vowels with the vowels contained in the samples, e.g. k → g, m → n, l → r;
(ii) recording vowels that are not in the text samples, e.g. h, k, p, t;
(iii) mistakes of association when the experimental persons associate a logatom in the samples with a meaningful word, which they could not understand clearly.

In the centre of most of the halls, the speech intelligibility is worse, due to that fact that less reflected sound arrives to these places besides the direct sound.

In the back rows of halls with rising rows of seats, intelligibility conditions are good.

For the different seat-groups as well as the speech intelligibility averages of the halls, great distribution values were obtained. It can be considerably reduced if the

number of persons taking part in the test is risen at least to 15 % of the seating capacity of the hall.

We can conclude that during speech intelligibility tests in each case the best results are with natural (live) speech (between 89,8-94,3 %), each hall falls to the subjective category of "good". The evaluation of the logatoms radiated through a loudspeaker gives results worse by 5-12 % (between 80,5-85,2 %) and according to it, each hall falls to the category of "adequate". With the conversion of the speech transmission indexes measured with the speech transmission meter, the PB word scores are between 70-75 % which values become worse by another 9-15 % but each hall belongs to the same category of "adequate".

It can be laid down as a fact that at all three measuring methods, changes in the results are similar according to the places (e.g. the values are worse in the centre of the halls).

The reverberation times in the examined lecture halls were rather different as it is shown in Table 2.

Table 2

Halls	seating capacity	m^3/person	T_{mid}(500-1000 Hz) (s) empty	occupied
BME E/A	221	6,1	1,90	$1,50^m$
BME E/B	408	5,6	2,20	$1,80^c$
BME KII/32	221	3,4	1,30	$1,05^m$
KMF I	156	5,2	1,80	$1,20^m$
BME KII/53	288	5,1	3,30	$2,70^c$

m=measured; c=calculated

The speech intelligibility results of the halls with rather different reverberation times (1,05-1,80 occupied) are close to one another except the extra reverberant BME KII/53 hall.

Conclusions

According to the measuring data we can conclude that the RASTI method gives good results, too, and the rating is very quick and easy. In our opinion, we can expect speech intelligibility values which are better by 5-10 % than the obtained results. We can conclude that with this equipment the changes in speech intelligibility could be measured quickly and well in the halls before and after their reconstruction.

It seems reasonable to modify the existing Hungarian intelligibility text samples, for example to create shorter samples with logatoms of three-syllables only.

We have evaluated speech intelligibility at the examined halls with logatoms of three-syllables of the test samples only. The dispersion of the data are a bit smaller and the results are worse by 1-2 %, i.e. they are more realistic.

According to our previous experience and our present

measurement results, we think that a value which is a little higher than the one in foreign recommendations for mid-frequency reverberation in occupied halls of small and medium size is also satisfying with peak values of 1,3-1,4 s. It has been also confirmed by our speech intelligibility tests. This possibility is very important mostly at the multi-purpose halls.

To increase the speech intelligibility in a hall is possible with simple room acoustic forming (e.g. with rising rows of seats and with sound reflecting surfaces placed near the lecturer).

References

- Hungarian Standards MSZ 3392-1954 ; MSZ 18153-1980
- Simon Gy.- Kotschy A.: 7th ICA, Vol.2. 19A 3, Budapest 1971
- Kotschy A.- Reis F.: 11th ICA, Vol.7. p.53-56, Paris 1983
- RASTI - Brüel & Kjaer Technical Review 3- 1985
- RASTI Measurements - Brüel & Kjaer Application Notes (BO 0116, 0123)
- Tarnóczy T.: Room Acoustics I-II. Akadémiai Kiadó, Budapest 1986.

ÜBER DIE STIMMBILDUNG BEI SCHLAGERSÄNGERN

Boglárka BALÁZS
HNO-Abteilung, Krankenhaus János, Budapest, Ungarn

In unserer Zeit gibt es immer mehr Schlagersänger und Schlager-Sängerinnen, und sie sind immer beliebter und populärer. Sie nehmen aber einen besonderen Platz ein unter den Künstlern, die von ihrer Stimme leben.

Schlagersänger geben nicht das Maximum ihrer erlernten Stimmbildung, sondern passen sich einem Trend an. Im klassischen Sinne des Wortes sind sie keine Sänger, denn sie müssen so singen, als würden sie sprechen; sie sind jedoch auch keine Schauspieler, denn sie müssen ja singen.

Um diesen Gegensatz zu verstehen wollen wir nun untersuchen, wie ein Schlager sein soll: melodisch, stark rhythmisiert, und er sollte nicht zu viele Töne enthalten, das heisst: leicht erlernbar sein. Er spricht den Menschen des Alltags an, und der Vortragende hat so zu singen, dass die Zuhörer leicht mitsummen können. Es darf nicht den Eindruck machen, dass es zum Schlagersingen einer besonderen Stimmaterie oder irgendwelcher Vorstudien bedarf.

Schlagersänger dürfen nur mit einem minimalen Vibrato singen, sonst wirken sie "operettenhaft". Sie singen also nicht, wie sie könnten, sondern wie es ihnen die Mode vorschreibt. Es ist kein Zufall, dass immer mehr heisere Beat- und Popsänger mit ihrer erkrankten Sprech- und Singstimme den Arzt aufsuchen.

Die von der Schwingung der Stimmbänder erzeugten Schallwellen verändern sich beträchtlich, ehe sie zu der Stimme werden, die unsere Ohren wahrnehmen können, und die der betreffeden Person eigen ist. Durch die Volumenveränderung der Luftröhre und der Bronchien kommt bei Aus- und Einatmung ein charakteristischer Luftdruck in der Subglottis zustande. Die Morgagnischen Ventrikel, Rachen, Zungen, weicher Gaumen, Nasenhöhle und Nebenhöhlen tragen ebenfalls zur Gestaltung des primären Kehlkopftons bei. Der Kehlkopfton verstärkt sich in den supraglottischen Räumen infolge der pharyngealen und der pharyngobuccalen Resonanz. Bei der Stimmbildung ist also schon ein durch das Ansatzrohr modifizierter Ton zu hören. Die für die Sprachlaute charakteristischen Formanten werden von den Lippen und der Mundhöhle gebildet -- während die musikalische Grundlage der Singstimme im supraglottischen und hypopharyngealen Raum geschaffen wird.

Bei einer richtigen Singtechnik geschieht folgendes: singt man höher, dann erhöht man die Grundfrequenz, das ist aber nicht alles: auch der anschliessende Resonanzraum muss sich anpassen. Im Prinzip liegt eine optimale Resonanz vor, wenn der Rachenresonator auf den tiefsten Oberton des Kehlkopftons, also auf den Grundton abgestimmt wird. Beim Singen ist ein spezieller Gebrauch der Resonanzräume erforderlich.

Was tut nun ein Schlagersänger? Will er die Tonhöhe ändern, dann will er zugleich das Vibrato vermeiden, also er gibt sich Mühe, die Resonanzräume nicht zu ändern, die Lippen- und Mundhöhlenformanten beizubehalten: er singt aus dem Hals.

Für den Phoniater bedeutet das folgendes: wenn die Veränderung des primären Kehlkopflautes nicht mit der entsprechenden Veränderung der einzelnen Resonanzräume im Ansatzrohr einhergeht, dann kommt es zu Störungen der Stimmbildung.

Die physiologische Form des Stimmeinsatzes ist der weiche Einsatz. Er kommt infolge einer feinen Abduktion der Stimmbänder zustande, dadurch

nämlich, dass die Stimmritze nicht zu dicht schliesst. Je grösser der Druck der ausströmenden Luft ist, desto intensiver schwingen die StimmbäNDER: Wenn die Stimmbänder beim Einsatz ganz dicht schliessen, bedarf es zur Öffnung der Stimmritze eines grösseren Luftdruckes als normal. In diesem Fall werden die Stimmbänder infolge der starken Spreizung und der Reibung, durch die mit hohem Druck ausströmenden Luft geschädigt. Das ist der harte Einsatz. Bei hartem Einsatz beträGT DIE Menge der Phonationsluft das doppelte derjenigen bei weichem Einsatz. Die Technik des Schlagersängers ist durch harten Stimmansatz gekennzeichnet.

Der Schlagersänger singt zwar ins Mikrophon, seine Stimme wird also verstärkt, trotzdem dürfen die achtzig bis hundert Dezibel nicht ausser acht gelassen werden, die er zu übertönen hat. Selbst wenn nur ein Soloklavier als Begleitung dient, gebraucht der Schlagersänger den harten Einsatz, da die Ohren des Publikums bereits daran gewöhnt sind. Hinzuzufügen ist, dass die Luft des Saales in der Regel voller Zigarettenrauch ist, und auch an den Tischen meistens kein stilles Publikum sitzt.

Wenn wir von Singstimme sprechen, müssen wir den Begriff des menschlichen Stimmregisters näher analysieren. Er wurde ursprünglich für die Orgel gebraucht. Wie Nadoleczny schreibt: "Unter Register verstehen wir eine Reihe von aufeinanderfolgenden gleichartigen Stimmklängen, die das musikalisch geübte Ohr von einer anderen sich daran anschliessenden Reihe ebenfalls unter sich gleichartiger Klänge an bestimmten Stellen abgrenzen kann. Ihr gleichartiger Klang ist durch ein bestimmtes konstantes Verhalten der Obertöne bedingt. Diesen Tonreihen entsprechen an Kopf, Hals und Brust bestimmte objektiv und subjektiv wahrnehmbare Vibrationsbezirke."

Wie entstehen nun diese Register?Durch Hebung des Kehlkopfes werden das Ansatzrohr kürzer, der Resonanzraum kleiner, die Klangfarbe heller. Durch Senkung des Kehlkopfes werden das Ansatzrohr länger, der Resonanzraum grösser, die Klangfarbe dunkler. Von den drei Registern - Brust-, Hals- und Kopfregister -- werden beim Sprechen in der Regel nur die Bruststimme und ein Teil der Halsstimme verwendet. Auch Schlagersänger singen nur im Brust- und Halsregister. Sie vermeiden das Kopfregister. Wenn man Kopfregister singt, wie z.B. Opernsänger, macht das stärkere Vibrato den gesungenen Text weniger verständlich. Diesen Preis bezahlt man in der Oper gern, weil hier die Schönheit der Stimme über allem steht. In einem drei Minuten langen Schlager ist der Text im gleichen Masse wichtig wie die Musik. Der Schlagersänger singt daher die gleiche Höhe um ein Register tiefer als ein Opernsänger, auch wenn seine Stimme nicht so schön klingt. So wird seine Stimme weniger schwungvoll, er muss also die Tonstärke durch Forcierung, eventuell durch Pressen steigern. Er singt die Töne an den Registergrenzen in dem tieferen, dem Sprechregister.

Der vielbeschäftigte, modeorientierte Schlagersänger befindet sich also in einer schweren Situation. Er beutet seine Stimmbänder aus. Das für das "Sängerknötchen" typische Luftsausen gibt ihnen sogar eine individuelle Färbung. Selbst dann, wenn der Schlagersänger keine Beschwerden hat, findet der untersuchende Phoniater Veränderungen an den Stimmbändern. Da ein Popsänger jedoch auch mit kranken Stimmbändern verhältnismässig lange singen kann, wird er wohl kaum zum Arzt gehen, im Gegensatz zu einem Opernsänger.

Wie kann nun ein Phoniater den Schlagersängern behilflich sein? Zunächst einmal darf er nicht versuchen, ohne die Mitarbeit eines Gesangpädagogen vorzugehen. Am wichtigsten ist, den Künstlern eine gute Atemtechnik beizubringen, damit sie die zum harten Einsatz nötige Luftmenge haben. Leider sind die Anforderungen an den Sänger nicht zu ändern so müssen wir versuchen, den Frequenzbereich zu finden, wo er mit einem Mindestmass an Anstrengung singen kann. Von der Höhe seiner Sprechstimme ausgehend

raten wir ihm, sein Repertoire auf nahe gelegene Töne zu transponieren.

Durch spezielle Übungen (Stimmhaltung und Stimmsteigerung) probieren wir die krankhafte und oft gut sichtbare Spannung der Kehlkopfmuskulatur und der Halsmuskeln abzuschaffen.

Mit speziellen Übungen kann man erreichen, dass Nasen- und Nebenhöhlenresonanz die Halsstimme als Kopfstimme klingen lässt. In manchen Fällen erzielt man schnelle und gute Erfolge durch Veränderung der Höhe der Sprechstimme; dabei bringen wir die alltägliche Sprechstimmhöhe an die Höhe der gewohnten Schlager des Sängers näher.

Die Kürze der Zeit erlaubt es mir nicht, die Möglichkeiten der Therapie ausführlicher zu behandeln, ich hoffe aber, dass es mir gelungen ist, einen Einblick in die shcwere und doch lächelnd verrichtete Arbeit der Schlagersänger zu geben. Und ich bitte meine Kollegen, die kranken Künstler, die sich an sie wenden, mit noch mehr Verständnis und Geduld zu behandeln.

Literatur

WENDLER, Jürgen--SEIDNER, Wolfram: Lehrbuch der Phoniatrie. Leipzig 1987.

AERODYNAMIC AND PSYCHOACOUSTIC PROPERTIES OF ESOPHAGEAL VOICE PRODUCTION

Tjeerd DE GRAAF, George L.J. NIEBOER and Harm K. SCHUTTE
Institute of Phonetic Sciences, Voice Research Lab, and
Centre for Voice, Speech and Language Disorders,
University of Groningen, Groningen, The Netherlands

Introduction

The university of Groningen has a long-standing tradition in the field of voice research, which was initiated by Jw. van den Berg who developed his Myoelastic-Aerodynamic theory of voice production in the 1950s. In this contribution, we want to report on a few recent developments related to the study of esophageal voice production, which also follow the Groningen tradition of Van den Berg, Damsté and Moolenaar-Bijl (1958). The aerodynamic and acoustic properties of the voices are compared with judgements on the related voice quality.

Measurements

The aerodynamic characteristics of the esophageal voice have been investigated in the Voice Research Laboratory and Centre for Voice, Speech and Language Disorders of the University of Groningen. Earlier the efficiency of normal and pathological voice production was studied by Schutte (1980), and in a research project in collaboration with the Groningen Institute of Phonetic Sciences his experience has been used with the simultaneous measurement of sound intensity, air flow rate and intra-esophageal pressure.

In the Groningen ENT-Clinic a tracheo-esophageal (TE) valve prosthesis has been developed (Nijdam et al., 1982), the so-called "Groningen Voice Prosthesis", and it has been the aim of our research project to gain insight into the physiological principles that are the basic elements for the production of the TE type and the injection type esophageal voice (the injection type voice [IE voice] is produced without a prosthesis). These aerodynamic properties are related to the phonetic quality of the voice and to the ability of the patient to communicate in a more or less acceptable way. The phonetic quality of the voices is determined by a series of perceptual evaluations with the help of techniques similar to those used by Boves and Van Herpt (cf. Boves, 1984).

In Figure 1 we illustrate the experimental setup for the measurement of various relevant parameters. After laryngectomy, the air passage to and from the lungs takes place through the tracheostoma, and there is no longer a direct connection with the pharynx. Pseudo-phonation with the esophageal sphincter is possible either by injecting air from the mouth during the realisation of plosive consonants (IE voice), or by introducing it through the valve prosthesis (TE voice), as indicated in the figure.

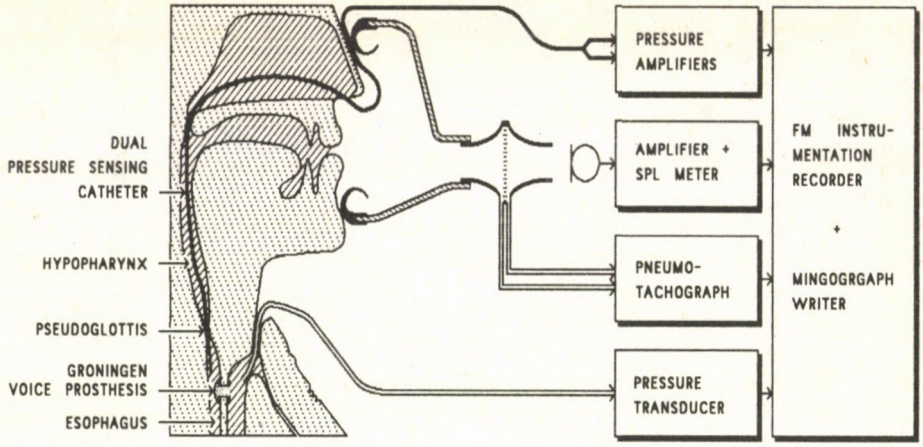


Figure 1: *Simultaneous registration of a number of pressure values, trans-pseudoglottal flow and sound intensity.*

The measurements consisted of the registration of sub- and supra-pseudoglottal pressure, sensed by two micro-tip pressure transducers inserted through the nose and positioned at a distance of 60 mm from each other at both sides of the pseudoglottis (De Graaf, Nieboer and Schutte, 1986). The catheter's diameter is 1.65 mm in the pseudoglottis, and it is assumed not to interfere substantially with phonation. The trans-pseudoglottal flow is determined with a Lilly flowhead, connected to a pneumotachograph; in addition, in patients using tracheo-esophageal speech, the intra-tracheal pressure is measured with an open catheter, connected to a pressure transducer. In the latter case we could obtain the trans-prosthesis pressure by subtracting the sub-pseudoglottal pressure from the intra-tracheal pressure. All measurements took place simultaneously with the registration of the audio signal and the sound intensity at a distance of 30 cm from the mouth. The patients were asked to pronounce /a/ and /i/ vowels and several consonant-vowel combinations with /a/.

A total of 496 measurements were performed on the phonations of 26 esophageal speakers, one group with IE voice, the other group with TE voice. Some of the speakers were able to produce both types of voice. The results for the mean values of some parameters are given in Table 1.

Aerodynamic variable	*All speakers*	*TE group*	*IE group*
p_{sub}	3.8 (2.2)	4.5 (2.4)	2.9 (1.7)
Flow (ml/s)	109 (93)	131 (112)	82 (62)
SPL (dB[A])	63.9 (9.1)	63.6 (8.7)	64.5 (9.7)

Table 1: *Mean values of sub-pseudoglottal pressure (p_{sub}), trans-pseudoglottal flow and sound pressure level (with standard deviations).*

The spread in these data is quite high, especially in the sub-pseudoglottal pressure

	Scales	*English translations*
1	zwak–krachtig	weak–powerful
2	onvast–vast	slack–firm
3	niet hees–hees	not breathy–breathy
4	eentonig–melodieus	monotonous–melodious
5	schel–diep	shrill–deep
6	traag–vlot	dragging–brisk
7	hortend–vloeiend	jerking–smooth-flowing
8	dof–helder	dull–clear
9	uitdrukkingsloos–expressief	expressionless–expressive
10	slecht verstaanbaar–goed verstaanbaar	not intelligible–intelligible
11	langzaam–snel	slow–quick
12	lelijk–mooi	ugly–beautiful
13	laag–hoog	low–high

Table 2: *The 13 semantic scales used in the voice evaluation experiment, with English translations.*

and the flow. The sound pressure level, on the other hand, has a rather small standard deviation, due to the generally small dynamic potential of these speakers.

When we consider the mean sub-pseudoglottal pressure, flow and SPL of these speakers, it becomes clear which variables are able to differentiate between the two groups of IE and TE speakers. Both sub-pseudoglottal pressure and flow are much higher in the case of TE speakers than in the case of IE speakers. The values of the SPL did not discriminate between both voice types: both reached values of about 67 dB at a distance of 30 cm.

Perceptual evaluation

The speech material was judged by a group of 85 listeners, both untrained judges and speech therapy students. They scored one minute of running speech of each speaker on 13 semantic 7-points scales. The speech samples, each with a duration of one minute were assembled from recordings of a number of sentences read aloud. In the perception experiment 13 contrasting Dutch terms were used for voice evaluation on the semantic scales shown in Table 2 (cf. Boves, 1980).

After factor analysis on the scores, three main factors turned out to be important: the "tempo"–factor, which comprises the scales 6, 7 and 11; the "voice quality"–factor, which is related to the scales 1, 2, 3, 4, 8, 9, 10 and 12, and finally the "pitch"–factor, which emerges from the scales 5 and 13.

The aim of our investigations was to correlate the aerodynamic data on the one hand and the evaluative data on the other. Throughout the calculations, a differentiation was made between the two groups of speakers mentioned. Emphasis was put on correlations between variables obtained during the aerodynamic measurements, and variables obtained from the rating experiment. The relationships between both sets of variables were investigated by means of product moment and multiple regression analy-

sis. The main questions in the analysis were: which aerodynamic factors are important for a voice that is judged to have good quality, and how well can voice perception be predicted from aerodynamic measurements.

Conclusions

From the results of the aerodynamic measurements and of the perceptual evaluation, we conclude:

1. The mean air flow correlates well with good voice quality;
2. Perceptual voice quality of TE speakers can be reliably predicted for 76% by measuring 4 aerodynamic variables: mean trans-pseudoglottal air flow, mean SPL, intra-tracheal pressure and prosthesis resistance;
3. Perceptual voice quality of IE-speakers is mainly determined by speech tempo or fluency;
4. Perceived pitch clearly correlated with measured SPL in the IE group. This is not the case in the TE group;
5. Correlation of perceptual voice quality with sub-pseudoglottal pressure is lower than expected.

References

1. BERG, Jw. van den, MOOLENAAR-BIJL, A.J., DAMSTÉ, P.H.: Oesophageal Speech. *Folia Phoniatrica,* 10. 1958, 65–84.
2. BOVES, L.: *The Phonetic Basis of Perceptual Ratings of Running Speech.* Foris Publications, Dordrecht, 1984.
3. DE GRAAF, T., NIEBOER, G.L.J. & SCHUTTE, H.K.: Production of Different Types of Esophageal Voice, Related to the Quality and the Intensity of the Sound Produced. *Folia Phoniatrica,* 38. 1986, 292.
4. NIEBOER, G.L.J., SCHUTTE, H.K. & DE GRAAF, T.: On the Reliability of the Intraoral Measuring of Subglottal Pressure. In: *Proceedings of the Tenth International Congress of Phonetic Sciences,* Foris Publications, Dordrecht, 1984, 367-370.
5. NIJDAM, H.F., ANNYAS, A.A., SCHUTTE, H.K., & LEEVER, H.: A new Prosthesis for Voice Rehabilitation after Laryngectomy. *Archives of Otorhinolaryngology,* 237. 1982, 27–33.
6. SCHUTTE, H.K.: *The Efficiency of Voice Production.* Dissertation, Groningen University. 1980, University of Groningen, The Netherlands.

ON THE PHONOLOGICAL RELEVANCE OF SOME NON-PHONOLOGICAL ELEMENTS

John KELLY
University of York UK

Underlying the theory of phonetics in a fundamental way are the two notions of 'posture' and 'gesture': and by virtue of this they have also influenced the ways in which many phonological theories have been constructed. The idea that speech is made up of a sequence of postures of the speech organs is of some antiquity and became well-established in the nineteenth century literature. The theory allowed that postures were linked one to another by movements of the organs of speech, but took it that the nature of such movements was preordained and, accordingly, predictable. The postures plus the linking movements constituted a gesture much of which was of no interest from the phonological point of view, and of relatively little from the phonetic. A lot of what happened could be left out of account in description and, accordingly, in analysis. Many classic texts urge that much observable phonetic detail should be discarded in this way at an early stage in the search for what is 'distinctive'. Phoneticians have commented that this 'posture'-based approach was not an accurate or illuminating account of what happens in speech; but this consideration was subordinated on the whole to another, namely, that the approach provided a succesful analytical tool, mapping neatly as it did onto a consequent 'segment' in the domain of phonology.

Over and above this, the introduction of the idea of 'broad distinctions into the study of sound systems focussed attention on a minimally specified element as the phonological prime, the minimal specification being applied to the 'postural' item held in isolation and viewed as a kind of oral (and aural) counterpart of the graphic letter. Specification of this kind was directly in terms of those physically present productive components that set a particular 'posture/segment' off 'distinctively' from others of its kind.

These developments meant that phonetic and phonological theory came to concern itself with only certain aspects of only certain points in the time and space continua of spoken language. And it was to these certain aspects and points in time that much instrumental work was applied.

The work presented here investigates components of the speech-complex that are not central to this 'posture/gesture' tradition. An earlier paper (2) makes reference to 'resonance' patterns associated with certain articulation in a number of accents of English. By resonance we mean here such things as the 'clearness' and 'darkness' of the phonetics text-books, typically mentioned in connection with, for English, the lateral consonant. The argument in (2) is that resonance effects of the type exemplified by 'dark l' co-occur with other articulations that make manifest the consonants of English; and that these resonance effects are of a long-domain nature. The present paper takes these investigations further and complements the corrobatory spectrographic material of (2) with findings derived from electropalatography.

This technique is described in (1). Like most others it has its limitations, and these must always be borne in mind. But valuable indications are to be had from it. In this paper we shall take the limitations as read and concentrate on a discussion of the electropalatographic findings.

As a preliminary to this something should be said about the place we

should assign to such finding. The work in question took, and takes, impressionistic (and necessarily subjective) techniques as being the first line of attack in work on speech of all kinds. 'Impressionistic' is to be taken in the widest sense here, as including visual and kinaesthetic as well as auditory judgements. Instrumental techniques are used to supplement these judgements. As they are in another mode, and quantifiable in ways in which the earlier judgements are not, there is no simple way in which instrumental records can be interpreted vis-a-vis subjective judgements. The artefactual nature of our equipment also blurs any picture that we might move towards creating for ourselves. In these less than ideal circumstances we must needs present instrumental results with a 'for what they are worth' caveat whilst trying to ensure that their ultimate usability is brought to a miximum by such precautions as are commonplace in science: replicability, statistical validity, and the like.

The kinds of phonetic description that we want to make are, then, firstly, the result of our turning our auditory and kinaesthetic capabilities on oursleves as speakers of English: and of later extending this observation to a number of other speakers. When some reasonably sound idea has been gained about what is happening in terms of articulatory and allied events, and when analytic records in the form of detailed transcriptions have been made, electropalatographic techniques are used to assess what, if anything, in the articulatory record might be interpreted as the counterparts in that mode of the auditory and kinaesthetic impressions.

The results presented here are perforce very preliminary and, to some degree, tentative. They relate to one subject and to relatively few occasions. They are presented as being of at least phonetic interest and of potentially phonological significance, both in the sense of mattering for the phonlogy of (this accent of) English and, if proven to be of general validity, for the construction of phonological theories.

During the discussion some of the impressionistic findings will be assembled and talked about informally; they will be accompanied by samples of relevant instrumental findings.

As has been said above, most references to resonance phenomena as understood here, have to do with laterals. (2) extends the range of the discussion to the English  articulations; and it is extended here to yet other phonetic items. To cover the range of these items the term 'resonance' is used rather freely. For dorsals, for instance, 'clear' and 'dark' resonance have to do with fronter or backer articulation respectively, whilst for apicals the phonetic events in question constitute a wider range of things, including secondary articulations as well as or instead of fronter or backer place of approximation. Over and above this, the ways in which resonance categories typically associate themselves with the set of apical items is different from the ways they do this with dorsals.

Almost invariably the phonetic phenomena associated with the presence of one or another resonance category have extent within the speech complex of a kind that overlaps changes of state on other, non-resonance, strands of this complex. They are best viewed, then, as <u>parametrically</u> associated with these.

'Clear' and 'dark' are labels that have not usually been listed amongst the phonological diacritica for English consonants. But in such cases as 'Why was <u>Henry late</u>?' as opposed to 'How did <u>Ken relate</u>?' there is, for my accent (usually called 'North-West Midlands') a distinct 'frontness' or 'clearness' of articulation in the second over the section marked  and more indeterminately so at ..., to set off against the 'darkness' of the first utterance. This is not to be put down simply to the presence of

any one phonetic item in these utterances (although it would be interesting if it were, given the extent of the resulting resonance). Rather it has to do with the occurence in the structure of these utterances of a number of things simultaneously, and of their ordering. A condideration of the words 'late' and 'relate' produced in isolation shows that the lateral is 'clearer' in the second than in the first, clearer, that is, when it is preceded by a ⅄ articulation in a preceding non-prominent syllable. In this accent of English the ⅄ articulation is typically clear, where the word 'typically' is related to a theory of acme articulation. When acting as an acme articulation, a phonetic item has its typical resonance value, which value it shares with neighbouring sounds. This means that it is not the case that all syllable-initial laterals will be clear. Their resonance characteristics are, rather, the result of other considerations, here that in the structural position in question ⅄ has acme function together with its typical value for resonance. The l of 'late' is in quite different circumstances. In 'grow' ⅄ is in different circumstances too: g is the acme items here and ⅄ is dark. In the utterances first considered above the presence of k and h in 'Ken' and 'Henry' are material too, since the first utterance is fronter (= clear) than the second in isolation.

What rules ordain acme function in phonetic complexes is not yet known. Certainly velars seem to have this function allocated to them regularly at the expense of other classes of items. We have seen this for 'grow' above, and it appears again, under quite different structural conditions, in 'It's Eric again' and 'It's Alec again'. In these utterances both ⅄ and l are clear, as is the velar articulation: these are to be compared with 'late' and 'relate' above and with 'It's Terry' and 'It's Telly' (2), in which ⅄, for this accent, is clear whilst l is dark, with, of course, the attendant extent implications.

These implications too will be subject to rules. What the rules are that govern the physical extent of resonance effects as they relate to the resonance structures of utterances is not yet known. But a beginning has been made. If we consider the utterances 'Barry came to my mind' and 'Ballet came to my mind' and the electropalatographic records that correspond to them we find that the velar contacts mappable onto 'c...' are appreciably fronter in the first of the two, where ⅄ has acme function. This is not the case for the corresponding portions of 'Tarrant came out' and 'Talent came out', produced with only velar closures at the '...t c...' place. These are quite different phonological structures, the second having a geminated, glottalised consonant absent from the first. Fig. 1 shows parts of the instrumental records that underlie these observations.

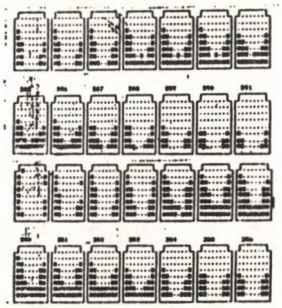


Figure 1. Velar contacts in 'Ballet came...' (upper) and 'Barry came...'

From the range of work we have done there are indications, additionally, that specifications along the resonance parameter vis-a-vis the others have to do with the manifestation of accent. The utterances 'This prank came to an end' and 'This plank came to an end' were produced with one prominent syllable each, 'prank' and 'plank'. The durations of the utterances were 185 and 180 csecs respectively. Two further utterances repeated these words, but with, this time, prominence at 'prank/plank', 'came' and 'end'. Durations in this case were 185 and 175 csecs overall for the two utterances. In the second pair the velar contact at '...ank c...' was fronter for 'prank' than for 'plank'; in the first case the contacts were identical. It is noteworthy here that we are dealing with a geminate again. This finding suggests that the provisional 'Geminate Rule' above needs to be modified to include a reference to particular accentual prominence configurations.

If resonance effects of the kind described here work in the way we suspect they do they are clearly of great importance as carriers of information about such things as syllable groupings within and across structural boundaries at other levels, accentual prominence, and rhythm (3). They may accordingly play a role for listeners in establishing syntactic and lexical identifications.

These elements of speech have had little publicity in the past, within the narrow approach to phonology outlined at the beginning of this paper. But we have no prima facie reasons for excluding them from our attention: and more recently a more liberal approach to phonology, consonant with the one we advocate here, has been urged, as in (4): '...listeners appear to rely on and be sensitive to all acoustic variations that a vocal tract can produce. If this is true, it may be useful for phonological and descriptive purposes to use a limited set of features, but it would not be valid if we want to understand real production and perception of speech. Rather than trying to reduce the number of features we should see how the multitude of features are used in different contexts for communciative purposes'. This non-reductive and polysystamic approach is one that our electropalatographic work, tentative as it is, takes as its guideline and goes some way to endorse. The study of the minutiae of utterance is not, in our view, a frivolous exercise, but rather the prerequisite to 'that total grasp of the phonic data on which all valid phonological and grammatical description - of whatever persuasion - must rest' (5).

References

1. HARDCASTLE, W.: New methods of profiling lingual-palatal contact patterns with electropalatography. Work in Progress, University of Reading Phonetics Laboratory 4. 1984, 1--40.
2. KELLY, J. & LOCAL, J.: Lond-domain resonance patterns in English, IEE Conference on Speech Input and Output, London, 1986.
3. LOCAL,J.: Some rhythm, resonance and quality variations in urban Tyneside speech. Studies in the pronunciation of English: forthcoming.
4. LÖFQVIST, A.: Review of McNeilage, P.: The production of speech, Phonetica 42, 1985, 157.
5. SHARP, A.: Stress and juncture in English. Transactions of the Philological Society. 1960, 104-135.

AVICENNA AND HUNGARIAN PHONETICIAN J. BUTTLER ON THE PECULIARITY OF AFFRICATE ARTICULATION

Alla PAVLOVA
G. V. Tsereteli Institute of Oriental Studies,
Georgian Academy of Sciences, Tbilisi, USSR

One wouldn't exaggerate by saying, that the number of authors come close to the number of definitions, given to the group of consonants, named affricates. Escaping the final definition, the phonetic structure of affricates raises a lot of questions before the researchers. We consider that some of these questions can find solution proceeding from peculiarity of the affricate articulation, described by Avicenna (1) in the XI century and experimentally analysed by J. Buttler (3) in the second half of the XX century. The subject here is concerned with the role that moisture (saliva) plays in the mouth cavity during formation of the affricates.

Inspite of the new apparatus used in experimental phonetics, in the second half of the XX century we can still find the definition of affricates as compound sounds which consists of the stop and the spirant (5). But even those, who consider this group of consonants as simple sounds, have different approach to revealing the stop and fricative elements (fases), their succession and correlation. Most part of the authors conceive these articulatory elements as successive (stop, then fricative), while L. Belgeri (2) marked the presence of both, the fricative and stop elements, from the begining till the end of sounding. Having investigated many languages and operated on a considerable material, L. Belgeri regarded the affricate as a "mixed" sound with undevided so- und curve, since some fricative elements appear in a segment which could be attributed to the stop component and vise versa.

Great experimental work on affricate research has been carried out by the czech phonetician B. Hálá (6;7). According to his data, the middle of the stop phase of czech affricates (c and č) is characterized by high assibilation degree. Transition from one phase to another is hardly discernable, that makes it almost impossible establish phase deviding point (7). Nevertheless, the scholar not only distinguished the two articulatory elements but also analysed the principle of their correlation in articulation of hissing and hushing affricates in one language, as well as correlation principle between the indentical affricates in different languages.

There are also different points of view on the explosion in affricate articulation. A number of the researchers (8;9;11) deny the presence of explosion on experimental basis. Not only the fact of explosion, but also its place in succesive raw of articulatory phases is the subject under discussions. Majority of the investigators mark the presence of explosion after stop phase, but B. Hálá (6) had a different approach to the problem. Analysing the oscillograms, he determined that explosion follows

not stop, but fricative element. He points out that this last phase was taken in account only by L. Belgeri and presented in his dissertation.

Spectral analysis of the czech language sounds, which was carried out by B. Borovičkova and V. Maláč (4), registered that the end of the stop component of c and č affricates is not always accompanied by explosion recognizable on the spectrogram. On the other hand, some false explosions appear in both affricates throughout fricative component and sometimes at the end of the latter. But, according to their view, this final "explosion" is not a real one. Such an "explosion", observed on spectrogram as a vertical line, may be the result of quick opening of aperture.

Data, obtained by us by means of the oscillographic analysis of the chinese language material, coincides with the observations of czech phoneticians. On our oscillograms the fricative fase of the affricates is characterized by the presence of 2-3 irregular explosive-like "splashes", at times with the absence of such an impulse at the end of stop component (11).

It must be mentioned, that cinema-x-ray research of L. Kovaleva (10) determined that "main feature of articulatory dynamics of the affricates is the presence of two, rarely three lingual and prelaryngal impulses... characterized as broken impulses".

In the view of the above stated, the phonetic treatise of Avicenna (1), containing the conceptions of scientist on the articulatory and acoustic phonetics, is of particular interest. Describing in detail the phormation of sounds in arabic, Avicenna speaks about the role of such physiological factor as saliva. According to his definition, the sounds of speech are differenciated by the character of articulatory surfaces, takint part in their formation: "... sometimes they can be soft, sometimes hard, dry or moist. Sometimes implosion occurs in moisture which cracks, then bursts spreading or elongating, or remaining at the same place" (1). Avicenna notes not only the presence or absence of moisture, but also the process which moisture undergoes during sound articulation.

Let us dwell on the role of moisture in articulation of arabic affricate ǧ (ǧim). Avicenna regards it as a simple sound occurring in a time interval between implosion and release. Formation of the affricate ǧ is represented by the scientist in the following way: complete implosion at first, followed by release resulting in a narrow aperture which permits the passage of air with hushing. Further on "hushing decreases and this causes the cracking of moisture pushed into space (between the teeth, - A.P.), then follows the explosion of moisture. This cracking doesn't spread width - wise, but occurs at the place of release". (1) As it is evident, the affricate is characterized by release, and not by explosion. And what B. Hálá considered as the consonant explosion accomplishing the articulation, was regarded by Avicenna as explosion of moisture, filling the aperture which was formed during the release of articulatory surfaces. Avicenna also mentions 3 sounds, resembling ǧ, which are absent in arabic. According to V. Akhvlediani (1) the question deals with the triad of hissing affricates. The role of moisture is again underlined: "Moisture is used similarly with that of

gim and it forms their sounding. This moisture is ready behind the closure..."

In the beginning of the sixties J. Buttler (2) having analysed a great amount of special literature, had defined the affricates as the consonants with the least clear phonetic characteristics. Trying to explain the mechanism of formation of affricates with simultaneous presence of stop and aperture, nomed by him vibrational, J. Buttler turned his attention to the physiological factor. He aimed at determining the role of moisture (saliva) in affricate articulation by means of atropin injections to the informator. The experiment showed that affricate oscillograms, recorded in normal conditions, are distinguished by more complicated (more tousled) sound curve. As researcher notes, this reflects an uneven distribution of energy in affricate sounding (one automatically makes an analogy with already mentioned false explosions, splashes...). As opposed to Avicenna, J. Buttler considers that the explosion of moisture starts the gradual release of the stop, but the latter can occur without such an explosion too.

According to J. Buttler, moisture, covering articulatory surfaces, plays a double role: it affects the gradual release of the stop, developing an aperture, while the viscosity of saliva connects the articulatory surfaces to a certain degree. It also could be understood as the particles of saliva fill up the newly formed aperture, which could be opened further on. This is the only explanation which could be given to the gradual release of the stop or the simultaneous presence of stop and aperture. Amount, distribution and viscosity of saliva are easily changeable, indefinite factors. That's why the extremely changeable affricate articulation is frequently observed. The scientist comes to the conclusion that saliva plays an undoubtable role in affricate articulation.

A number of our data, obtained from the chinese affricate oscillograms, could be considered as the confirmation to J. Buttler's conceptions on the affricate articulation.

References

1. AKHVLEDIANI, V.: Avicenna's phonetic treatise. Tbilisi, 1966 (in russian).
2. BELGERI, L.: Les affriquées en Italien et dans les autres principales langues européenes. Institut de Phonétique de L'Université de Grenoble, 1929.
3. BUTTLER, J.: The formation and acoustic Structure of Affricates. Acta Linguistica t. XIV, F. 3--4. Budapest, 1964. 263--275.
4. BOROVIČKOVA, B.--MALÁČ V.: The spectral analysis of czech sound combinations. Rozpravy Československé Akademie Ved, Rada Společenskyh Ved. Roč. 77, Seš. 14. Praha, 1967.
5. DELATTRE, P.: Studies in French and comparative Phonetics. London--Paris, 1966.
6. HÁJA, B.: Une contribution à l'éclaircissement de la nature phonétique des affriquées. Zeitschrift für Phonetik und allgemeine Sprachwissenschaft, HT. 1/2. Berlin, 1952. 77--93.

7. HÁLA, B.: La nature des consonnes mi-occlusives mise en lumière au moyen des procédés experimentaux modernes. Actes du X-ieme Congrès Internationale de Linguistique et Philologie Romanes. Strassbourg, 1962. Paris, 1965. 887--897.
8. HACHATRIAN, A.: On the affricates of armenian language. Proceedings of the Armenian Ac. of Scien. 11. Erevan, 1962 (in russian).
9. HACHATRIAN, A.--AIRAPETIAN, V.: Experimental research of consonantal phonemes in armenian literary language. Erevan, 1971 (in russian).
10. KOVALIOVA, L.: Affricate articulatory dynamics in speech. Dissert. abstr. Kiev, 1981 (in russian).
11. PAVLOVA, A.: Oscillographic research of the chinese affricates. Proceedings of the Georgian Ac. of Scien. Matsne 4. Tbilisi, 1978. 155--172. (in russian).
12. ZINDER, L.: General phonetics. Moscow, 1979.

CHARAKTERISIERUNG DER PHONETIK DER GESPROCHENEN SPRACHE IM HEUTIGEN RUSSISCHEN

Erzsébet RÉPÁSI
Lehrstuhl für russische Sprache und Literatur, Pädagogische Hochschule 'Bessenyei György', Nyíregyháza, Ungarn

Die Literatursprache /Schriftsprache oder Hochsprache/ ist der Sprachgebrauch der Schriftlichkeit der Menschen von Bildung, anders ausgedrückt: sie ist die Sprache der Schöngeistigen Literatur, der hochwertigen Presserzeugnisse und der öffentlichen Reden. Sie ist die anspruchsvollste sprachliche Variante von hohem Niveau, die als ein Ideal hisichtlich der Phonetik, Morphologie, Lexik und der Satzlehre anerkannt werden kann. Diese ideale Variante der Sprache, die in eine in den Grammatiken festgelegte Norm hat, wird in der den Normen vollkommen entsprechenden Form nur sehr selten realisiert. Sie tritt meistens in einer weniger anspruchsvollen sprachlichen Variante auf, die aber über der Sprache der Menschen verschiedener Berufe, unterschiedener gesellschaftlicher Schichten bzw. über den Dialekten steht. Diese Variante der Sprache, "das am häufigsten benutzte Verständigungsmittel des Alltags und die Erscheinungsform der Sprache, in der heute am stärksten sprachliche Neuerungen und Entwicklungen vor sich gehen"[1] bezeichnet man mit dem Wort Umgangssprache. "Die Umgangssprache ist ... der lebendige, gesprochene, der dialektischen Elemente bare, mit Norm geregelte Sprachgebrauch der sprachlich geschulten, gebildeten Menschen. Das ist die nicht vorbereitete, nicht geschriebene, vorher nicht geschriebene, vorher nicht formulierte, spontane, ungebundene

Tätigkeit der Denkarbeit und der Formulierung des Textes ohne nachträgliche Kontrolle. Diese Tätigkeit wird durch bestimmte Normen geregelt. Diese Variante wird gewöhnlich bei der persönlichen oder offiziellen Kommunikation. zu Diskussionen und Außerungen gebraucht."[2] Die hier zitierte Definition charakterisiert -- meiner Meinung nach -- aber nicht nur die Umgangssprache, sonder auch die gesprochene Sprache. Umgangssprache und gesprochene Sprache sollten voneinander getrennt werden. "Gesprochene Sprache liegt dann vor, wenn die sprachliche Äußerung vom Sprecher in spontaner Rede oder in freier Manuskriptrede im Hinblick auf den Hörer hervorgebracht wird und wenn der Hörer präsent ist, die Hervorbringung der Äußerung unmittelbar verfolgen kann, denselben Situationsbezug wie der Sprecher hat und Rezeption durch Reaktion anzeigen kann."[3] Die gesprochene Sprache kann mit den von den kodifizierten Normen der Literatursprache unterschiedenen Abweichungen charakterisiert werden. Diese Abweichungen werden durch die objektive, soziale, gesellschaftliche und kommunikative Mikrosituation, aber auch durch die kommunikative Rolle beeinflußt. Stilistisch lassen sich folgende Kategorien dieser Sprachäußerung unterscheiden: a/ Stil der spontanen oder ungebundenen Rede, b/ Stil der Vorlesung, c/ Stil der reproduktiven oder interpretativen Rede, d/ Stil der annähernd reproduktiven Rede.[4]

Das von mir untersuchte Material ist eine in Anbetracht der Intonation angefertigte Umschrift russischer Texte, die in der Variante der gesprochenen Sprache formuliert worden sind.[5] Unter den obenerwähnten Kategorien können sie in die erste /"spontane, ungebundene Rede"/ einklassiert werden. Zu

den Themen der Dialoge gehören das kulturelle Leben /Bücher, Theater, Kino, Musik, Wohnungsfragen, geselliges Leben, Sport, Kinderzeit usw./. Es handelt sich um die Dialoge von Bekannten, Freundinnen, Familienmitgliedern. In Kenntnis dieser Umstände kann ohne besondere Analyse vorhergesagt werden, daß die Aussprachеweise "lockerer", das Sprachtempo schneller ist als in der Literatursprache. Wir werden wahrscheinlich mehrere Reduktionen finden. Die genaue Analyse der sprachlichen Erscheinungen, die hier -- wegen der strengen Beschränkung des Umfanges dieser Abhandlung -- ausführlich nicht erörtert werden kann, aber in meinem Referat zur Zeit der Konferenz genauer dargelegt wird, macht es möglich, daß wir die wichtigsten Merkmale der russischen gesprochenen Sprache auf dem Gebiet der Phonetik klassifizieren und dadurch die Charakterzüge der russischen Sprache genauer beschreiben können.

G.A. Barinova, die die phonetischen Charakterzüge der gesprochenen russischen Sprache bisher vielleicht am gründlichsten analysiert hat,[6] unterscheidet acht wichtige Erscheinungen. In dem von mir untersuchten Material zu drei typischen Charakterzügen kann man neue Angaben finden:

1. Zu der quantitativen Reduktion der Vokale.

Hierher kann der Ausfall bestimmter Vokale in unbetonter Lage eingereiht werden. Das Maß der Reduktion des Vokals hängt von der Position des gegebenen Lautes ab. Diese Position bedeutet einerseits die Stelle des Lautes in der Wortform, andererseits aber auch seine Anpassung an die ganze Intonationsstruktur des Satzes.

2. Zur Reduktion, wobei eine Silbe ausfällt.

Die Folge der vollen quantitativen Reduktion der Vokale ist der Ausfall von ganzen Silben und außerdem die Konsonantenhäufung, die in der Aussprache noch weiter vereinfacht wird.

3. Zur Vereinfachung von Konsonantengruppen.

Diese Erscheinung kann vor allem in einer intervokalen Position untersucht werden. In der gesprochenen Sprache ist eine allgemein wirkende Tendenz, daß die Konsonanten in der Aussprache oft ausfallen oder verschwächt werden.

Alle, die die russische Sprache als Fremdsprache erlernen wollen, sollten nicht bestrebt sein, die gesprochene Sprache nachzuahmen. Es genügt, sie zu verstehen. Es ist aber leichter, wenn wir ihre wichtigsten Charakterzüge kennen.

Literaturverzeichnis

1. Hans, Berthold: Zur Syntax ostmitteldeutscher monologischer gehobener Umgangssprache. Wissenschaftliche Zeitschrift der Universität Rostock -- 18. Jhg. 1969. Gesellschafts- und sprachwissenschaftliche Reihe, Heft 6/7: 561-4, 561.
2. Penavin, Olga: Nyelvjárás és köznyelv. Nyelvművelő Füzetek. Fórum Könyvkiadó. Ujvidék, 1986. 23.
3. Hans, Berthold: Theoretische Fragen der Beschreibung gesprochener Sprache. Wissenschaftliche Zeitschrift der Pädagogischen Hochschule "Clara Zetkin" Leipzig III/1984: 8-9.
4. Wacha Imre: A beszéd hangzásának stílusa. /Szövegfonetika./ Fejezet a kiejtési kézikönyvhöz. H.n., 1974.1-4. Zitiert von Vértes O. András: Bevezetés a magyar hangstilisztikába. Nyelvtudományi értekezések 124.sz. Akadémiai Kiadó, Budapest, 1987. 5.
5. E. A. Zemskaja i L. A. Kapanadze (Otv. red.): Russkaja razgovornaja reč. Teksty. Nauka, Moskva, 1978, 27-80.
6. E. A. Zemskaja (Otv. red.): Russkaja razgovornaja reč. Glava II. Fonetika. Nauka, Moskva, 1973, 40-151.

SYLLABIC CONSONANTS IN COMMON KARTVELIAN

Rusudan ASATIANI
Language Typology Department, Institute
of Oriental Studies, Academy of Sciences
of Georgian SSR, Tbilisi, USSR

The phonological system of the Common Kartvelian language consists of three types of phonemes (2):

Vowels - phonemes, which form syllables in any position;

Consonants - phonemes, which do not form syllables in any position;

Sonants - phonemes, which form syllables in some positions.

Reconstruction of sonant phonemes - /r̥/, /l̥/, /m̥/, /n̥/, /i̯/, /u̯/ in Common Kartvelian becomes possible through the anaysis of correspondences of Georgian, Čanian-Megrelian (Zanian) and Svanian reflexes:

*ʒ₁ᵃɣ̊ᵉ > ʒᵃɣᵉ (K) : ʒᵒɣᵒr (Z) : žᵉɣᵂ (Sv)

*c₁xr̥a > cxra : čxoro : čxara

*ʒ₁m̥a > ʒma : žuma : žum-ie

*č̣ᵢnᵉx > sa-c̣nex : — : č̣inax

So: *S > S(K) : VS(Z) : VS(Sv).

Analogous reflexes occure in the following examples:

A.		B.	
tba : toba/ tub		dɣe : dɣa : deɣ	
gb(oba) : gib : ab// b		qba : - : qab	

So: C C(K) : VC(Z) : VC(Sv)

Here, in case A, the development of anaptixic vowel is meant surpassing accessive complexes (4, 113); e.g. *t̥b > *t̥ᵒb > tob/ tib; in case B - the vowel metathesis is meant, which can be explained by the strong tendency of Svanian to form closed syllables (4); e.g. * dɣe > *dᵒɣe > *deɣe > deɣ.

American linguist Alan Bell (1, B3) considers that "unless there is evidence for the status of the associated vocalic element as a segment, the consonant should be specified as syllabic". According to such functional specifications, in cases A and B, the existance of syllabic phonemes (in this case - non-sonorants) in Common Kartvelian may be supposed.

Such an explanation seems to be more likely as basically like events might be explained by the general theoretic assumption: Common Kartvelian has syllabic phonemes (sonants, fricatives, stops).

Such an assumption is typologically quite acceptable — Alan Bell describes a type of languages, where potentially any phoneme can be syllabic. This assumption requires the modification of the Common Kartvelian phonological system.

Looking for traces of syllabic phonemes within consonant groups seems quite appropriate. Various clusters are being reconstructed in Common Kartvelian:

A		A'	B		B'
bɣ	px pq	zɣ sx	bg	pk pk̇	s k̇
dɣ	tx t̥q	s₁x	dg	tk t̥k̇	s₁k̇
ʒɣ	cx c̣q		ʒg	ck c̣k̇	
ʒ₁ɣ	c₁x c̣₁q		ʒg	c₁k c̣₁k̇	
žɣ	čx č̣q		žg	čk č̣k̇	

Let's consider the reflexes of these harmonic clusters in Kartvelian languages. The voiceless consonant clusters of group A are easily reconstructed on the Common Kartvelian level (4, 81). As for the voiced ones, some differences for the Common Kartvelian and Common Georgian - Zanian levels are to be seen.

On the level of Georgian-Zanian clusters with voiced consonants may be reconstructed just as easily as clusters with voiceless consonants are, but on the Common Kartvelian level, reconstruction of such clusters becomes more complicated, because in Svanian either there are no suitable reflexes, or even if they occur, they are of the C_1VC_2 type:

*dɣ > dɣe(K):deɣ(Sv)
*ʒɣ > ʒɣv : ǯoɣv
ʒvan: žoɣvan

Thus, it can be assumed, that on the Common Kartvelian level the reconstruction of voiced type A harmonic clusters cannot be accomplished: bɣ, dɣ, ʒɣ, ǯɣ -groups in Svanian are separated by a vocalic element of an indefinite timbre and therefore, according to Alan Bell, in these positions the existance of syllabic phonemes may be supposed. We give priority to the reconstruction of */d̥/ and */ʒ̥/, for generally, in the Kartvelian languages (in the case of syllabic sonant (2))#-C position is strong, while C-V position is weaker and reveals different structures according to areals.

We have more problems in reconstructing of harmonic clusters type B on the Common Kartvelian level: in Svanian their reflexes are less numerous and they are of the C_1VC_2 or $C_1,C_2//C_2$ type; C_1,C_2 be voiced as well as aspirated; e.g.:

*dg > dgam(K): gem//gm(Sv)
dg : lhg
*pk > pkvil-i: pek
*tk > tkum-a: li-ku-isg

In the case of existance of C_1VC_2 reflexes, C_1(=*/p/) may be retained. In other cases, where the loss of C_1 causes prolongation of the infinitive-forming vowel (4): $C_1,C_2 > C_1,C_2$ $(C_1)VC_2 > (V)C_2$ - syllabic */d̥/ and */t̥/ may be supposed. The analysis of type B harmonic complexes suggested the possibility of reconstructing aspirated syllabic phonemes: */t̥/ and */p̥/. One of Alan Bell's basic typologic requirements: "If a language possesses voiceless syllabic stops, then it possesses voiced syllabic stops" - is observed here, as we have already reconstructed the voiced syllabic stop */d̥/ and such an opportunity exists also for the */p̥/ (see below). The analysis of A' and B' type clusters reveals analogous changes:

*zɣ > zɣva(K): zɣua(Z): ʒuɣua(Sv); So */ʒ̥/
*sx > si-sx-l: zisxis(M):dicxir(Č):zisx(Sv);
*s_1x > sxva:šxva: - : ešxu; So */s_1̥/;
*s_1k > skvinča:kvinča:kvinči(Č);
*sk̥ > ska: sk̥: (m)ska, (m)ka(Č).

Besides harmonic complexes of type A and B there also exist the so called "quasi-harmonic" non-accessive clusters, which cannot be reconstructed on the Common Georgian-Zanian level in case if one of the consonants is voiced:

*bǯ > bǯ:biǯg; so */b̥/
*bz > bzu:buz; so */b̥/
*bǫķ > bǫķvn:biǫķon, biskon; so */b̥/
*(S)ʒ > rʒe:bža, bǯa:lǫǯe; so */ʒ₁/

But on the Common Kartvelian level, even if both consonants are voiceless:

*ps > ps:ps:li - s-ēr; so */s̥/
*sç > gasçreba:wrapa:li-sr-e; so */s̥/

Reconstruction of the accessive, disharmonic clusters for the Common Kartvelian level seems problematic for any consonant (whether voiced or voiceless) (2):

*kb > ķb:ķib:ķib -; */b̥/
*tb > ţb:ţub,ţib:ţib:ţub; */b̥/
*ġb > gb:gub:ġib:ǯab/ǯb; */g̥/
*qb > qb: - : - : qab; */b̥/
*kb > čb/ceb: - : - :kab/kb; */b̥/
*cd > cd:čod:čod: - ; */d̥/
*šd > - : šķid: - šged/šgd; */d̥/
*gs > qs(gs): rš:š:ǯiš//ǯš; */s̥₁/
*qd > xd:rt:xt,xt:qad/qed/qid; */d̥/
*kd > cd:č:č:kad/kd; */d̥/
*gz > gz:z,rz:(n)gz,z:līzi; */z̥/
*qc₁ > xc:rč:x /kč: - ; */c̥₁/
*qs > xs :xš,(r)sx:(r)cx,(m)cx: - ; */s̥/
*qs₁ > xs:š(ǫ):š(u): - ; */s̥₁/
*ţp > ţb:ţub:ţib:teb/ţb; */ţ/
*ķs₁ > ķs:š:(n):š:usg/ušk */s̥₁/

In reconstructing syllabic consonants we keep the following principles:

1. Fricative surpasses stop in syllabic formation;
2. Voiced > voiceless;
3. # -C position > c-v position

The arrangement of these principles has hierarchical power: 1> 2>3.

Syllabic phonemes in Zanian undergo the following changes:

$$*C_1VC_2 > *C_1{}^{\circ}C_2 > *C_1(r,n)C_2 // *C_1C_2(ǫ) > (r,n)C_2 // C_2(ǫ).$$

Thus we can suppose that the phonological system of Common Kartvelian consists of two kinds of phonemes:

Vowels - phonemes, which form syllables in any position;
Syllabics - phonemes, which form syllables in definite positions.

The system of syllabic phonemes has the following structure :

Sonants: */r̥/, */l̥/, */m̥/, */n̥/, */w/, */j/
Fricatives: */z̥/, */s̥/, */s̥₁/
Stops: */b̥/ */p̥/
*/d̥/ */ţ̥/ */ţ/
*/ʒ₁/ */c̥₁/
*/g̥/

Such representation of syllabic phonemes supposes the following stages in the development of the Common Kartvelian languages:

I. CVCV type structures (see (2), (5));
II. Loss of the segmental status of vowels; formation of consonant groups; formation of syllabic consonants (compare

with Alan Bell's so called III-type language);

III. The Common Kartvelian stage: existance of syllabic sonants, fricatives and voiced stops; syllabicity of voiceless stops and back consonants become weak (they occur only with type B harmonic clusters);

IV. The Common Georgian-Zanian stage: existance of syllabic sonants; distributional restrictions of syllabic stops and fricatives (they are possible only with accessive, non-harmonic consonant sequences);

V. The Modern stage: loss of syllabic phonemes.

References:

1. BELL,A.: Syllabic Consonants, Working Papers on Language Universals, Stanford, California, November, 1970;
2. GAMKRELIDZE,T., MACHAVARIANI,G.: The System of Sonants and Ablaut in Kartvelian Languages, Tbilisi, 1965 (in Georgian);
3. KLIMOV,G.: Etimological Dictionary of Kartvelian Languages, Moskow, 1964 (in Russian);
4. MACHAVARIANI,G.: The Consonant System of Common Kartvelian, Tbilisi, 1965 (in Georgian);
5. CHIKOBAVA,Arn.: The Oldest Structure of Roots in Kartvelian, Tbilisi, 1942 (in Georgian).

EARLY SECOND LANGUAGE CONTACT - ACQUISITION OR LEARNING?

Leslie BARRATT
Indiana State University,
Terre Haute

Ilona KASSAI
Linguistics Institute,
Hungarian Academy of Sciences, Budapest

Introduction

The two terms 'acquisition' and 'learning' have traditionally been used in the literature to delineate two separate processes with respect to languages. The first term has been applied to the unconscious way children develop competence in their first language while the second has been reserved to the conscious way in which adults master a second language. In between one can find numerous other viewpoints on the issue (1, 2, 3). This paper offers a modest contribution by describing the strategies of a child exposed to a second language in the pre-school years.

Background

The child in this study is one of the authors' daugther, Elissa. She was 3;9 when she arrived in Budapest for a 10-month stay. Before going to Hungary E. had reached an adult-like stage of language acquisition in English. She had been an early talker, using telegraphic speech by 18 months. She had no command of a second language, although she had occasionally heard other languages, most often Dutch, spoken around her. Before leaving for Hungary, she learned the Hungarian expressions for the following: No, yes, Watch out!, That's not allowed!, Where's the bathroom?, boy, girl, mother, father, color terms and numbers to ten. She arrived in Budapest on September 2, 1987 and on September 15 she started in a Hungarian pre-school where none of the teachers or children spoke any English. E. attended school 5 days per week from approximately 9 am until 4 pm until June 13, 1988, the last day of school. She also attended a gym class twice a week after school, taught in Hungarian. By December, E.'s accent was described as 'native' by Hungarian speakers. By March, her Hungarian was good enough to pass her off as Hungarian. She could carry on full conversations with other children or adults without anyone knowing she was American. Outside of school, E. was exposed to both English and Hungarian.

Data and results

Most of the data described here were collected on June 4, 1988. A 2-hour recording was made with E. playing with two Hungarian sisters, Lilla, age 5;7 and Virág, age 3;0. Those tapes revealed that, although in everyday conversations E. sounded to all Hungarians like a Hungarian, her speech was not identical to that of other Hungarian children of her age. Specifically, she sounded at 4;6 like a child about one year younger because she made native speaker errors that children generally stop making by the time they are her age. Types of the native child errors E. made are listed below.

NATIVE ERRORS IN HUNGARIAN

S y n t a x

word order and agreement - Miért rajta van ezek? instead of Miért vannak rajta ezek? 'Why are these on her (the doll)?' (here the verb should be second and in the plural); Meg akarom nézni ezt a babákat. instead of ezeket

'I want to look at these dolls' (the demonstrative pronoun should be in the plural);

P h o n o l o g y / M o r p h o n o l o g y

Suprasegmental errors

intonation - in yes/no questions final syllable rise/half fall generalized from 2-syllable questions to 3-syllable and more than 3-syllable questions, where adults use rise/fall extending on the last two syllables. E.g. Te tudsz mindent angolulba, Lilla? 'Do you know everything in English, Lilla?' (rise/half fall on the syllable -ba instead of rise/fall on -lulba);

stress - utterance-final extra stress in WH-questions, e.g. Hol a babaszekrény? (with stress on both the WH-word and the last syllable) 'Where is the doll's closet?';

segmental errors

consonant harmony - galangommal for galambommal 'with my pigeon', ne bobálj for ne dobálj 'don't throw!';

definite article - a óra instead of az óra 'the clock' (although the a/az distinction is exactly the same as the English a/an distinction, she doesn't make these errors in English with a/an;

underextraction - bábo should be báb 'puppet' (nominative derived from accusative, by attaching the linking vowel to the stem);

underanalysis - nyakájába instead of nyakába 'onto her neck' (double possessive marking due to the underanalysis of the genitive nyaka; úszoda for uszoda 'swimming pool' (by analogy with the verb úszni);

regularized irregular forms - hosszúbb instead of hosszabb 'longer' (failure to produce irregular comparative);

S e m a n t i c s

mindig 'always' for soha 'never' in negative sentences.

These examples thus illustrate that E. made many of the errors Hungarian children make in the acquisition of their native language. These acquisitional errors were not surprising, given E.'s age and the nature of her language contact. However, the data also revealed other errors, ones which Hungarian children do not make, but which adult learners of Hungarian might make. These are given below:

NON-NATIVE ERRORS IN HUNGARIAN

S y n t a x

agreement - overuse of the copular verb van - Ugye, hogy milyen szép van? instead of Ugye, hogy milyen szép? 'Isn't it pretty?' (the copular verb must not appear on the surface, the intonation is also faulty);

flectional suffixes missing - Én ismerek egy Lilla, amelyik hülye Lilla. (first Lilla is missing the object marker -t) 'I know a Lilla who is a yucky Lilla'; Mindenki, aki kör kívül van. instead of körön kívül (locative suffix missing) 'Everybody outside the circle'; Nekünk is van eper. instead of eprünk (double marking rule is not applied) 'We also have strawberries'; Az, aki kell tapsolni for akinek (dative suffix is omitted) 'The one who must clap her hands';

P h o n o l o g y/M o r p h o n o l o g y

aspirated stops at word beginnings - tapsolni, kanál;

ø/,/y/ uncertainties - she doesn't always hit the target, especially in long sentences, in non-initial position;

reversal of the members of the short and long vowel pair ɛ/e: - felesegé instead of felesége 'his wife', ehés instead of éhes 'hungry';

M o r p h o l o g y

vowel harmony - Barbimnek instead of Barbimnak, 'to my Barbie', csütörtöken for csütörtökön 'on Thursday' (here only rounding harmony is lacking, fronting harmony is observed), óvodánél for óvodánál 'at the kindergarten';

suffix redundancy - angolulba for angolul 'in English', estébe for este 'in the evening' (-ba/-be is the equivalent of in);
incomplete rule application - úszjál instead of ússzál 'swim' (non-application of the assimilation rule obligatory for stems ending with strident), szalvétam for szalvétám 'my napkin';
misordering and wrong choice of bound morphemes - babámakkal for babáimmal 'with my dolls';
simogatottad - past tense of definite verb modelled on that of indefinite verb.

S e m a n t i c s

az egész Barbit for összes Barbit 'all Barbies'.

One observation which is immediately apparent is that most of E.'s native-like and non-native errors in Hungarian were in the areas of phonology and morphology rather than in syntax. Another obvious generalization is that many of E.'s non-native Hungarian errors are the type which traditionally would have been described as i n t e r f e r e n c e errors from English. Some of her non-native errors, however, are clearly not from English. The vowel harmony errors and some of the underanalysis errors cannot be classified as interference so they might be considered as resulting from learning strategies, evidenced e.g. by frequent self-corrections. To sum up, Elissa clearly has relied on universal strategies of language acquisition in order to learn Hungarian. We can also see that she invokes her knowledge of English in her strategies with regard to Hungarian. However, some of the errors suggest that E. also relies on an analytic strategy different from that of a child acquiring a first language. This strategy produces non-native type errors, but ones which do not have English as a point of departure. For example, all of E.'s vowel harmony errors involve the vowel /ɛ/. She seems to have realized that this vowel has something of a neutral status, but she has overgeneralized this status so that she uses /ɛ/ on many suffixes which are supposed to alternate.

Having shown that E. depends on several strategies in Hungarian, it seems logical to ask if her contact with Hungarian and in particular her use of new strategies have had any effect on her acquisition of English. An analysis of her errors in English suggests that they have. We noted earlier that E. was in the adult-like stage of English language acquisition. The kinds of errors she is still making with English which might be considered acquisitional, come from modals (e.g. We should might do it.) and tag questions (positive tag with positive statement, negative tag with negative statement) in syntax; from /θ/ and /ð/ phonemes in phonology; from comparatives and superlatives (e.g. Breakfast is the importantest meal.), irregular verbs (e.g. sawn for seen, derivational suffixes (e.g. undust the furniture, productive use of -ish) in morphology. After her contact with Hungarian, E. started to make the kinds of errors which cannot be considered as native. Some of them are given below.

S y n t a x

calques - You know that. for You know. (from Hungarian Tudod. with definite ending), You can take JUST as much as you want to. (from Vehetsz, amennyit CSAK akarsz.), Get it if you know to! (in a keep-away game, Fogd meg, ha tudod!), confusion of he and she (from Hungarian ő);
word order - There are they. (in response to Where are they? modelled on the Hungarian order: Ott vannak. and Hol vannak?), On my lap she should go. (Az ölembe üljön!);
prepositions missing or incorrect - I'm afraid from the dog. (a kutyától 'dog+from', Thank you the necklace. (from Hungarian Köszönöm a nyakláncot. 'necklace+acc., Can you take it down? (for 'off' referring to getting gum off a pen - Le tudod venni?), negation - Until I do NOT go. (Amíg NEM

megyek.);
P h o n o l o g y
intonation - Hungarian intonation, especially in calling
segment - trill used for English flap, e.g. in Daddy.
S e m a n t i c s
Turn it up! (a window in a car) for Roll it up! (Hungarian Csavard fel!).

As becomes obvious, before going to Hungary almost all of E.'s acquisitional errors in English concerned morphology. This parallels her situation in Hungarian. Interestingly, however, after being in Hungary for ten months, most of her problems with English came from syntax, an area which had not previously given her trouble in English and did not give her trouble in Hungarian. The predominance of morphological errors in E's English before language contact may be explained so that this level of language was the most complex for her. As, however, morphology in English is far less rich than in Hungarian, she started to master Hungarian morphology with almost the same tabula rasa as Hungarian children do when acquiring their native language. As for her "post-contact" situation in English, we suggest the following interpretation. Going back to America she has no more problems with morphology as relying on her extended knowledge of Hungarian morphology she is able to resolve all problems raised by English morphology. On the other hand, her problems with syntax might come from the fact that the strict word order of English appears to be less convenient for expressing her communicative intentions than the free or, anyhow, more flexible order of Hungarian.

Discussion

On the basis of the results of error analyses both in E.'s Hungarian and English performance, the question asked in the title of our paper cannot be answered with either one of the alternatives. Rather, both have to be used, substituting and for or. With respect to E.'s Hungarian competence, we can state that discovery procedures characteristic of the acquisition process are parallelled by the knowledge of a language already acquired as well as the analytical approach characteristic of the learning process. On the other hand, E.'s strategies have changed with respect to her native language as a result of the language contact. She seems to be more conscious of language than before, more actively involved in the acquisition process. We could interpret her new attitude so that thanks to her knowledge of the Hungarian language, there are competing linguistic structures at her disposal for the expression of the same reality. She then must make her choice. While doing this, she unconsciously might think over the advantages and disadvantages of these competing structures regardless whether they pertain to Hungarian or English.

On reflection, the distinction between acquisition and learning seems to be a highly sophisticated one.

References

1. Krashen, S. D.: Second Language Acquisition and Second Language Learning. Oxford, 1981.
2. Krashen, S. D.: Principles and Practice in Second Language Acquisition. New Jersey, 1987.
3. Hakuta, K.: Mirror of Language. The Debate on Bilingualism. New York, 1986.

A QUASI-PALEONTOLOGICAL APPROACH TO SPEECH
(Mythos and Logos under the Pretext of Aristotle)

Lóránt BENCZE
Department of Modern Hungarian
Loránd Eötvös University, Budapest, Hungary

As an **introductory remark** I would like to point out that in our efforts to interpret the world around us and in us and to share this knowledge with one another, we can feel the world (in both senses of the verb) and can be satisfied or dissatisfied with it. Such a knowledge is highly coloured by our emotions and instincts and is expressed in 'mythos'. It is, however, "hardly suitable for us when seeking the truth" (1). Therefore, we also set up 'logos', i.e. a system of categorized scientific knowledge (2), which " is a reasonable way for accounting for facts" (1). Whenever we crystallize the first way by means of the second we create 'mythologos' -- poetry, arts and religion. Yet, even when we concentrate on the second way we are unable to abandon the first, except in Utopias, which have failed and have been inhuman when practiced in the history of mankind. "Knowledge once thought to be absolute, indubitable, is now seen as provisional or even probabilistic... as it is socially justified belief... and located in the community of practitioners" (5). This is why combining the two ways seems to be the most adequate and most common procedure for us.

Etymological evidence proves that the ways of the 'mythos' and the 'logos' are coded in the primordial metalanguage, at least in Indo-European languages (and may have been in every language). In particular I refer to the verbs concerning speech activity, which were divided into eight etymological groups by Buck and can be divided into two main types. **Type 1**: i m i t a t i o n / s i m i l a r i t y of sound (e.g. Greek mythos, mytheomai,etc., English mutter, French mot,etc.; Greek spharageo /meaning 'crackle', 'sputter', 'hiss'/, Danish sprage/'crackle'/, English speak, etc.), or of visual contact ('point out', i.e. denote by means of seeing an object). **Type 2**: synonyms for the r a t i o n a l phenomenon of c a t e g o r i z a t i o n (quantity and discretion) meaning 'arrange', 'order', 'assembly', 'select' (e.g. Greek lego /'say'/, Latin lego/'read', originally 'pick out', 'select', 'collect', 'count', 'recount', 'enumerate', 'tell', etc.) (4).

In Hungarian, grouping is difficult both for the uncertainties in etymology and for the various sources of borrowings (e.g. olvas /'read'/, originally meant 'count', 'enumerate', 'tell'/;- it is either of Finno-Ugric or more probably of old Turkish/Chuwash/ origin; mond /'tell', 'say'/ is perhaps either Finno-Ugric or Indo-European /cf. Latin 'monere'-'admonish', 'foretell'/; szó/szól 'word/say' meant 'hymn', 'prayer' in Finno-Ugric languages but was 'report', 'speech' in old Turkish (9).

I will call Type 1 mythos-type language and Type 2 logos-type. They directly relate to the two ways of approaching and interpreting the world mentioned above. In the **history** of the humanities -- and mainly **of linguistics** -- this opposition has had various names, more or less with the same meanings. The following is a list in which the opposition may not always be consistent:

1. Mythos-type	2. Logos-type
physei	thesei
natural	artificial
iconic	arbitrary
concrete	abstract
intuitive	logical
emotional	rational
free	systematic
artistic/poetic, magic, mythic/	scientific
continuous	contiguous
analogue	digital

(cf. the different functions of cerebral hemispheres)

An **evolutionary viewpoint** asks whether "in the beginning" was the logos or the mythos? It is surely pointless to try to decide whether the mythos or the logos came first. There are no sound or video tapes recording the birth of hominids several million years ago. Yet, it will not be fruitless to discuss the matter as it relates to the present state of affairs. Adam and Eve may have become homo sapiens by eating the apple of knowledge and thus acquiring the logos-type language and thinking in addition to mythos-type feeling and expressing themselves (3)(6). In them, mankind was driven from the garden of Eden into this world of good and evil, true and false, beautiful and ugly. In Type 1 we have kept the memory of Paradise lost. In Type 2 we have gained the ability to discern. In order to survive as an individual, Type 2 became predominant in all men; in order to survive as a species, Type 1 remained predominant -- mainly -- in women. Possessing an exclusively Type 1 knowledge we die of hunger; possessing exclusively Type 2 knowledge we die of ruthless wars. Type 1 provides us with a holistic view of the world and of human beings which also includes feelings and instincts, enabling us to be merciful and compassionate and to create poetry, arts and religion. Type 2 gives us the power to be the lords and rulers of the world, to categorize, to reason and give evidence, and thus cultivate sciences and to develop technology.

On the one hand, in the **phylogenesis** of language we are inclined to give priority to iconicity /visuality/ based on development stages like activity --> imitating activity by pretending /visual!/ --> imitating activity with audio-visual presign --> imitating activity with merely audible sign --> imitating activity by means of arbitrary signs. Distorted sound imitation can become arbitrary as in the Hungarian word csecsemő 'baby' [tʃ] [tʃ] [m] -- which reminds us of the sounds of sucking (7).

On the other hand, and at present, iconic looks secondary and complicated in relation to arbitrariness. Similarity is usually based on semantics and not on sound, e.g. boat reminds us of ship above all and not of boot (6). The difference in onomatopoeic words in various languages shows that they possess a certain degree of abstraction. Nevertheless, no word within a language community is arbitrary for the community (11). Many parents have found that certain children, especially girls, can imitate sounds very early without making proper distinctions among phonemes and structures. Others, mainly boys, start speaking comparatively later, and from the beginning almost perfectly with discretion and structure.

Thus, the double coding (8) is present in many ways and in various degrees of iconicity and arbitrariness (cf. image, diagram, metaphor /Peirce/, motivation /Saussure/, isomorphism, onomatopoeia, etc.). Coinciding with this static presence a permanent dynamism of the two can also be discovered (cf. remotivation, demotivation, erosion, etc.)(8)(10).

In Type 1 we get to know the continuity of the world. We can fix a value to every moment of reality. This is then an analogue system. At the same time, we have a highly developed digital system in our thinking and language. In this way Type 2 measures values in discrete units. The values are integral multiples of one another, although they can be arbitrary in their refinement. Through Type 2 we are confronted with the contiguity of the world, i.e. sequence of quantums, discrete units, or 'things' as we usually say. Neither type exists in language and thinking in its pure form. For example the digital system is not exclusive in telling a story (with arbitrary signs, i.e. words), but is accompanied by the analogue system as we give an account of events in a sequence as they took place. The most sophisticated integration of the two systems is metaphor which makes up the essence both in poetry and in scientific invention and discovery (3).

The **conclusion** we may draw from this is that the three approaches to knowledge (Type 1, Type 2 plus the combination of the two) can be looked at as imitation (mimesis) of the world in our mind. Though the combination of the two types is the most common and most adequate for man (3), there is always a temptation to consider them antagonistic, to see a gap mainly between Types 1 and 2, to play off one against the other, to think that the question is to believe **or** to prove. Our behaviour is not sine fundamento in re, not without reason, when we see an "inherent contradiction" here (7). It is buried in the double layer of language built on language with a permanent two-directional movement between the two. This dynamism of language is the sine qua non of our lethal adjustment to the changing world.

REFERENCES

1. Aristotle, Meteorologica. With an English Transl. by H.D.P.Lee. London-Cambridge, Mass. 1962. 356 b, 357 a.
2. Bencze, L., "Mythos and Logos: Aristotelian Synonyms?" Annales Univ. Scient. Budapest. de R. Eötvös nom.S.Lingu. (1986) XVII. 165-8.
3. Bencze, L., On the Metaphorical Animal. ibid. (1981) XII. 215-22.
4. Buck, C.D., Words of Speaking and Saying in the Indo-European languages. American J. of Philology. (1915) XXXVI. 1-18.
5. Davis, Ph.J., Applied Mathematics as Social Contract. ZDM 88/1. 10-14.
6. Décsy, Gy., Sprachherkunftsforschung. Bd. II. Berlin, etc. 1981. 8,20,10.
7. Fónagy, I., The Languages within Language: Toward a Paleontological Approach of Verbal Communication. In: Approaches to Language. Anthropological Issues. W.C. McCormack and S.A. Wurm eds., The Hague-Paris, 1978. 79-134.
8. Fónagy, I., Double Coding in Speech. Semiotica 3, (1971), 189-222.
9. Lakó, G., ed. A magyar szókészlet finnugor elemei.1-3. Budapest, 1967-1978.
10. Haiman, J., Natural Syntax. Iconicity and Erosion. Cambridge, 1985.
11. Verhaar, J.W.M., On Iconicity and Hierarchy. Studies in Language. 9/1. (1985), 21-76.

UNITÉ DU SYSTEME ET CLASSEMENT DE L'ENSEMBLE DES PHONÈMES /DANS UN BUT PÉDAGOGIQUE/

Ildikó BODNÁR
Mikszáth Kálmán Gimnázium, Pásztó

En étudiant les représentations du vocalisme et du consonantisme de différentes langues, une divergence importante apparaît au premier coup d'oeil. /L'auteur entend par représentations surtout les tableaux traditionnels, figurant dans les livres de grammaire et dans les manuels de phonétique de nos jours./

Les représentations du vocalisme sont remarquablement unifiées, celles du consonantisme apparaissent dans une très grande mesure dissemblables. En ce qui concerne la vocalisme, c'est presque toujours le système de D. Jones - son célèbre quadrilatère, accepté aussi par l'Association Phonétique Internationale - qui sert comme base pour la plupart des représentations actuelles. Ce quadrilatère qui peut arranger les voyelles de n'importe quelle langue, est le résultat d'une longue évolution.

Sans vouloir analyser la voie historique des classifications et des représentations des sons articulés, il faut pourtant mentionner le nom de C.F. Hellwag qui - en 1781 - avait donné pour la première fois une représentation triangulaire pour les voyelles allemandes. Bien que les systèmes triangulaires - et leurs continuations, les systèmes quadrilatères - reposent sur l'articulation, de nombreuses nouvelles recherches prouvent que la classification acoustique de même que le classement perceptif peuvent être présentés dans les mêmes cadres. Comme J. Herman écrit: "En raison des liens entre les mouvements articulatoires et les propriétés acoustiques des voyelles, ce fameux triangle vocalique est pratiquement identique à la disposition des voyelles dans un système de coordonnées logarithmique, où l'une des coordonnées porte les valeurs des F_1, l'autre celles des F_2."

D'une manière singulière, la représentation correcte des consonnes ne préoccupait pas trop les phonéticiens. Les tableaux de consonnes sont privés de la logique interne des systèmes de voyelles, qui, tout en présentant les voyelles d'une certain langue, reflètent aussi leurs interdépendances et en plus quelques "fonctionnements" de la langue vivante, comme les substitutions phonétiques, modifications dialectales, changements historiques.

Les tableaux actuels des consonnes, avec leurs variations innombrables, semblent être accidentels. Il est difficile de justifier les raisons de la place particulière des occlusives parmi les colonnes des modes d'articulation. L'ordre des autres colonnes des modes d'articulation et partiellement l'ordre des lignes des lieux d'articulation semblent être aussi le fait du hasard. C'est peut être la force de la tradition qui a laissé inchangé l'ancien classement grec où P-T-K se trouvaient en tête de la disposition tabulaire.

Mais si dans le domaine des voyelles cette représentation logique, à la fois articulatoire, acoustique et perceptive existe, pourquoi ne pourrait-elle pas exister dans le domaine des consonnes? Plusieurs auteurs, comme par exemple R. Jakobson en parlent: "Ainsi les différences entre quatre types de consonnes /vélaires, palatales, dentales, labiales/ se réduisent en fait aux deux oppositions de qualités phonologiques, que nous

venons de définir au point de vue de la phonation et que nous allons examiner maintenant du point de vue acoustique. Les consonnes postérieures s'opposent aux antérieures correspondantes par un plus haut degré de perceptibilité, souvent accompagné ceteris paribus d'un plus haut degré de durée."

En examinant un grand nombre de tableaux consonantiques il ressortait, que ces tableaux ne font qu'énumerer les consonnes de la langue donnée, sans montrer les rapports mutuels entre les consonnes et surtout sans les montrer entre les sous-groupes des consonnes comme occlusives, fricatives, nasales d'une part et labiales, dentales, palatales etc. d l'autre.

Une réorganisation du système des consonnes peut amener à la reconnaissance de ces relations existantes. La nouvelle représentation que je vais présenter reflète alors les mêmes interdépendances, les mêmes fonctionnements qu'on a trouvés dans le domaine des voyelles.

Il faut dire que cette question purement thérique n'est apparue qu'après les premiers pas pratiques de la réorganisation du système consonantique, dont le but était: rendre visible les rapports plus profonds, montrer quelques processus phonétiques des consonnes de la langue hongroise comme les différents types d'assimilation au cours de l'enseignement.

Si la parenté des systèmes des voyelles et des consonnes /plus correctement des deux sous-systèmes/ est aussi grande que l'on suppose, ces deux sous-systemes - selon les espérances de l'auteur - donneront un système unique. Le système des consonnes doit être alors "ouvert" vers les voyelles. La tentative de la représentation unique - présentée en détail au cours de l'exposé - est le résultat d'un travail de recherche de plusieurs années. Après avoir inventé que les deux systèmes se complètent vraiment, il a fallu trouver les justifications de toute la conception.

C'est surtout la linguistique historique traditionnelle et une nouvelle branche de la linguistique moderne, la phonologie naturelle /Natural Phonology/ qui ont fourni les arguments les plus précieux, les exemples les plus clairs. En dehors des exemples séparés de changements phonétiques, c'est surtout "l'échelle de la fortification" qui soutient les conceptions de l'auteur.

Les changements montrent une grande régularité, les processus phonétiques ont leur stricte logique. Le nouveau classement, le nouveau système unique intègre toute un série des résultats modernes. On montre par exemple l'affaiblissement comme un processus lent et se composant de petits pas. La même représentation de petits pas /ou de séries de changements graduels/ est possible dans le système nouveau. Les tendances montrent dans la même direction sur l'échelle de la fortification et chez l'auteur. /La réorganisation rend possible la présentation - et interprétation - de la plupart des changements phonétiques comme changements graduels./

On trouve quelques détails logiques même dans les systèmes anciens illogiques. Mais les relations qui étaient évidentes à l'interieur d'une colonne /colonne des fricatives p.e./, ne l'étaient pas entre les colonnes. On ne voyait pas les grands rapports entre les classes de modes

d'articulation, ni entre celles de lieux d'articulation. L'ouverture vers les voyelles dont il a été question dans la première partie est un bont exemple de relations plus complexes, de relations au niveau du système.

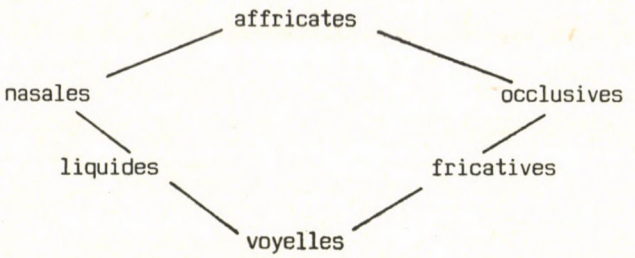
L'ordre logique, du point de vue articulatoire, acoustique et perceptif des consonnes est le suivant:

a) Les colonnes des modes d'articulation se rangent ainsi: 1. fricatives, 2. occlusives, 3. affricates, 4. nasales, 5. latérales, 6. vibrantes.
b) Les lignes des lieux d'articulation diffèrent moins de l'ordre traditionelle, il s'agit seulement de la transposition des labiales. /La justification aura lieu pendant l'exposé./

z	d	d͡z	n	l	r
s	t	t͡s			
ʒ		d͡ʒ			
ʃ		t͡ʃ			
j		J	ɲ		
ç		C			
ɣ	g		ŋ		
x	k				
h					
β	b		m		
f	p				
v					

Ce système consonantique est ouvert vers les voyelles des deux cotés. On connaît depuis longtemps que les fricatives /surtout j et β / sont très proches des voyelles. Nous savons aujourd'hui que les nasales et les

et les liquides - d'un point de vue acoustique - ressemblent beaucoup aux voyelles. La représentation spatiale peut bien faire voir cette situation:



Ouvrages cités:

1. HERMAN, J.: Phonétique et phonologie du français contemporain.
2. JAKOBSON, R.: Le classement phonologique des consonnes in.: Selected Writings I.

SYLLABIC CONSONANTS IN POLISH CASUAL SPEECH

Wolfgang U. DRESSLER, Liliana MADELSKA
Institut für Sprachwissenschaft
der Universität Wien, Vienna, Austria

1. In this contribiution we want to present some results of a phonogical (and phonetic) investigation of Polish fast/casual speech, where segmental and prosodic phonology interact. The analysis is done in terms of Natural Phonology (13, 6), which has been a framework for the study of fast/casual speech in many languages, e.g. (7, 8, 10). The primitives of this theory are phonemes (sound intentions) and phonological processes which both govern the inventory and phonotactics of phonemes and transform them into phonetic realizations. Most of these latter processes are backgrounding (or lenition) processes.

Usually phonologists investigate syllabic consonants (particularily obstruents) only if they are phonemes. The best survey is (2). The "reluctance to recognize that obstruents, too, may be syllabic has never been overcome" [(2), p. 155]. Similar to Russian (1), Polish has no syllabic consonants in either phonemic representation or in prescriptive formal speech [pace (5)]. In Polish casual speech, and even in the rather formal speech of newscasters syllabic consonants occur fairly regularly. So far, they have been investigated without the basis of ample corpora of spontaneous speech (3, 5, 11).

2. Our investigation is based on quality recordings of spontaneous conversations with 30 students made by the second author, and on phonetic transcriptions made by a team at A. Mickiewicz University, Poznań. Incomprehensible or mumbled fragments were excluded here, although they were taken into account in (9). All of the transcriptions (more than 60000 word tokens) have been organized into a computational "Dictionary of phonetic realizations", "Table of phonemes and their realizations" etc. (9). Consonants were transcribed as syllabic, when the transcribers agreed among themselves. Then spectrograms were made of all crucial and problematic instances: syllabic consonants proved to differ from their non-syllabic counterparts by longer duration (without a trace of the deleted vowel), as measured on spectrograms, and by higher amplitudes in their characteristic frequency ranges. Spectrograms clearly showed temporal compensation for the lost vowel by lengthening the syllabic consonant, and also compensation in intensity.

In the investigated texts, apart from other phonological processes, the deletion of consonants was far more frequent than the deletion of vowels , e.g. [t]-deletion applied to over 8% of all occurrences of phonologically intended /t/, [f]-deletion to over 6%; as to oral vowels, the rate of deletion was as follows: [a], [e] and [i]: 0.2%, [o]: 0.6%, [u]: 1.0%, and [ɨ]: 1.6%.

3. Less than one third of vowel deletions provoked the syllabification of a preceding or a following consonant. The high vowel [ɨ] - the least sonorous among Polish vowels - most easily both deletes, and then provokes the syllabification of adjacent consonants. Phonological backgrounding leading to vowel deletion is gradual: first vowel contrasts are reduced, then vowels delete, with syllabification of an adjacent consonant, and finally the syllabic peak is totally lost. (Further deletions are also possible).

4. Polish consonants can become syllabic according to the following conditions:

a. If the vowel is deleted, only a tautosyllabic consonant can become syllabic. Examples like Rubach's (11, p.73) [fɛ$d̪ɛ$ral$n̪ɛj]→ [fɛd̪ral̪n̪ɛj] "federal", [mɛ$lɔd̪ja] →[m̩lɔd̪ja] "melody" neither occur in our data nor were accepted by native speakers. Possible pronunciations might be rather [fɛd̪ral̪n̪ɛj, f̩d̪eral̪n̪ɛj, m̩lɔd̪ja], cf. [v ɔgu$lɛ] → [vɔg̩lɛ/vɔglɛ, *vɔg̩lɛ] "in general".

b. Only consonants adjacent to a deleted vowel can become syllabic, e.g. [mʃɨ$tsɛ]→ [mʃ̩tsɛ, *m̩ʃtsɛ]. If the adjacent consonant is itself deleted, the next consonant becomes a candidate for syllabicity, e.g. [zd̪ɛtsɨd̪ɔvawɛm]→ [z̩tsɨd̪ɔvawɛm] "I have decided".

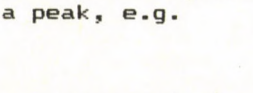
c. If there are two tautosyllabic adjacent consonants in a closed syllable, the more sonorous one will become syllabic. This is evident for sonorants vs obstruents, e.g. [bɨm] → [bm̩] "would", [t̪ɨlkɔ] → [t̪l̩kɔ] "only". In [juʃ] → [ʃ̩] "already", /j/ was deleted before vowel deletion. As to other sonority distinctions, our data are insufficient, but nasals seem to be more "sonorous" than /r/ on this account, e.g. [ɔ$ran̪$ʒa$d̪a] → [ɔrn̪ʒad̪a] "orange juice" (11, p. 74).

d. If adjacent consonants are in the same sonority class, the left-hand consonant becomes syllabic, e.g. [fʃɨstsɨ → fʃ̩stsɨ, fʃɨst̪kɔ → fʃ̩skɔ/fʃ̩st̪kɔ, sɨst̪ɛmat̪ɨʧn̪ɨ → s̩st̪ɛmat̪ɨʧn̪ɨ] "all (masc. pl.), all (n. sg.), systematic".

5. These conditions find the following explanation within Natural Phonology:

a. If a vowel is deleted and the prosodic syllable structure remains, no heterosyllabic consonants can be mapped to the syllable peak - unless resyllabification occurs in still faster/more casual speech.

b. The syllable consists of syllable rise towards the syllable peak (the highest position) and of syllable fall starting with a peak, e.g.



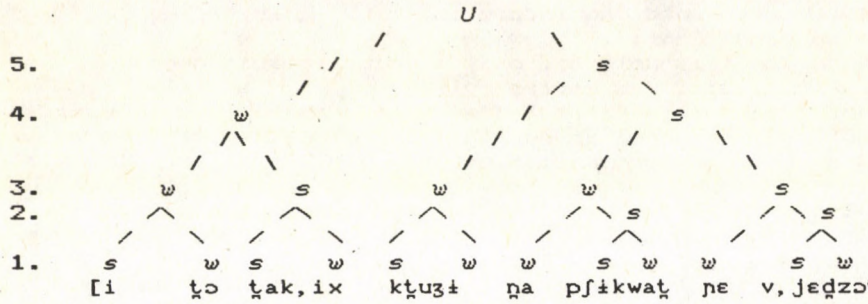
as in [fʃɨst̪$kɔ].

If the vowel is deleted, then only the adjacent consonants C2, C3 are candidates for peak assigment, because they are higher up than the more distant C1, C4.

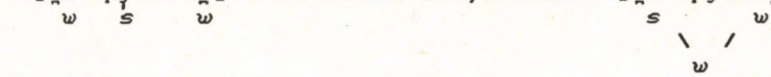
c. A more sonorous consonant is closer to the vowel peak than a less sonorous one, e.g. Therefore it is higher and a better candidate for peak assignment.

d. Comparable consonants are prosodically stronger in the syllable rise than in the syllable fall.
6. Desyllabification of syllabic consonants can be only studied within the rhythmic structure (prosodic tree) of a phonological phrase/sentence. Let us give an illustrative example, the following piece of formal speech

<i to takich którzy na przykład nie wiedzą>
"and that such (persons) who for example don't know".



Here we have 5 hierarchical levels of prosodic (stress) strength. In very slow speech only the lower stress levels 1-3 are actualized, i.e. there are four independent stress units [i ʈɔ ˈʈak,ix] [ˈkʈuʒɨ] [ɳa ˈpʃɨkwaʈ] [ɲɛ ˈv,jɛɖzɔw̃].
7. In moderately faster/more casual speech the hierarchical stress levels 4 and 5 are added. As a consequence of combining four independent stress units into a single one the actually audible stress distinctions are flattened, i.e. strong (s) nodes dominated by weak (w) nodes are weakened. Under this condition the vowel of the fourth (potentially) strong (s) syllable can be deleted, because it is dominated by a higher weak (w) node; since the syllable peak remains, it is reassigned to the preceding sibilant according to 4a, b: [ɳa pʃɨkwaʈ] → [ɳa pʃkwaʈ] "for example".
8. In still faster/more casual speech - of course particularly in long prosodic stretches - the hierarchical structure is flattened. As a consequence syllabic peaks can be lost with resulting resyllabification. In our example trisyllabic ..[ɳa $pʃ$kwaʈ]..can become disyllabic ..[ɳa $pʃkwaʈ]..



where the difference between potentially strong and weak is hardly audible; but there is indipendent evidence for assigning (s) to [ɳa], cf. <na to> "on that", <na nim> "on him", where [ɳa] has the (potential) stress (optional in [ɳa v,jɛɕ] "to a village").
9. Finally we would like to mention the relation between production and perception. When the tapes were played to native speakers, they noticed syllabic consonants (and vowel deletions) only if the prosodic structure was changed, but not if the number of syllables and syllable peaks was maintained, i.e. when listening they did not differentiate trisyllabic [ɳa

pʃkwaʦ] from trisyllabic [n̥a pʃɨkwaʦ]. However, when the recordings contained resyllabified, i.e. disyllabic [n̥a pʃkwaʦ], then they would comment on this sloppy pronunciation. In terms of Natural Phonology, they perceived according to both the segmental phonemic intentions (which always contained the vowel /ɨ/ of [pʃɨkwaʦ, pʃkwaʦ, pʃkwaʦ]) and the prosodic intentions, i.e. disyllabicity vs trisyllabicity. Therefore pronunciations in 4a, such as [mɫɔd̥ja] proposed by Rubach (11) would evoke something like an intended (non-existing word) /mlɨɔd̥ja/. Therefore such a pronunciation immediately arouses curiosity among native listeners, whereas [m̥lɔd̥, ja] does not, provided that it is embedded into a prosodically adequate utterance.

10. In this specimen analysis of Polish syllabic consonants we have tried to examplify the approach of Natural Phonology to the study of fast/casual speech. Time and space limitations have impeded us from providing more on the theoretical model and presuppositions of our approach.

References

1. BARINOVA, G.A.: Редукция гласхых в разговорхой речи. *In* S. Vysotskij, et.al., Развитие фонетики современного русского языка. Moscow, Nauka. 1971, 97--116.
2. BELL, A.: Syllabic consonants. *In* J. Greenberg (ed.), Universals of Human Language. Vol. 2. Phonology. Standford: Standford Univ. Press. 1978, 153--201.
3. BIEDRZYCKI, L.:Analiza fonologiczna polskich i angielskich spółgłosek nosowych i samogłosek. Unpublished Ph.D. diss. Warszawa, Warsaw University, 1971.
4. BIEDRZYCKI, L.: Samogłoski bezdźwięczne w języku polskim. Logopedia XIII. 1975, 14--24.
5. BOGUSŁAWSKI, A.: Sylaba a system fonologiczny. Prace Filologiczne XXXII. 1985, 59--65.
6. DRESSLER, W.U.: Explaining Natural Phonology. Phonology Yearbook I. 1984, 29--51.
7. DRESSLER, W.U.--SIPTÁR, P., Towards a Natural Phonology of Hungarian. (to appear, in Budapest).
8. DRESSLER, W.U.--WODAK, R., Sociophonological methods in the study of sociolinguistic variation in Viennese German. Language in Society 11. 1982, 339--370.
9. MADELSKA, L.: Mowa spontaniczna. Analiza wariantywnosci fonetycznej w wymowie studentów Uniwersytetu im. A. Mickiewicza. Unpublished PhD diss. Poznań, A. Mickiewicz University, 1987.
10. MOOSMÜLLER, S.: Sociophonology. *In* P. Auer -- A.di Luzio (eds) Convergence and Variation. Berlin, de Gruyter. 1988, 75--92.
11. PERLIN, J.-- MADELSKA, L.: Syllabification of consonants as a consequence of vowel deletion in fast/casual speech in the light of Polish facts. (to appear, in Warsaw).
12. RUBACH J. Changes of consonants in English and Polish. A generative account. Wrocław etc., Ossolineum, 1977.
13. STAMPE, D. A dissertation on Natural Phonology. New York, Garland, 1980.

DETERMINANTS OF SPECTRAL VARIATION IN SPONTANEOUS SPEECH

Olle ENGSTRAND and Diana KRULL
Institute of Linguistics
University of Stockholm, Stockholm, Sweden

Introduction

The experimental search for physical and linguistic determinants of phonetic variation typically involves the use of well-structured, minimal speech samples. Clearly, this is a necessary and fruitful experimental methodology. Ultimately, however, the insights gained from such "laboratory speech" data will have to be validated within the broader range of speech styles used in natural communicative situations. It is reasonable to assume that the systematical relationships between variables frequently observed in conventional laboratory experiments will show up also in spontaneous speech. In the latter case, however, this systematicity will probably turn out to be less transparent. The reason is that the phonetic shape of spontaneously produced utterances is typically influenced by several variables which, by definition, are out of the experimenter's control. On the other hand, in exploring the extent and nature of phonetic systematicity in spontaneous speech, we may well come across phenomena that have previously escaped our attention precisely because we normally tend to set up highly constrained experiments. The study of spontaneous speech may thus be a heuristic undertaking leading to new hypotheses that, in turn, may be subjected to more rigorous experimentation.

In a current research project[1] we are concerned with several phonetic aspects of the production and perception of spontaneous speech. This particular paper exemplifies an attempt to examine such speech in search for the various sources of spectral variation in vowels. The primary purpose is to find out whether such a variation can at all be predicted as a consequence of segment duration and, if so, to what extent. Part of the background of this experiment is also a series of papers (1,2,4,6,7) on spectral variation in vowels, starting with Lindblom's original reduction study (4). In Lindblom's study, vowel reduction was expressed in terms of formant "undershoot" relative to invariant acoustic targets. The amount of undershoot was determined by segment duration. However, subsequent research has demonstrated that durational and spectral properties of vowels may vary independently and that the full range of spectral variation can not therefore be sufficiently explained by a duration-dependent model (1,2,6,7). For example, Lindblom and Moon (6) presented evidence of durational-spectral independence in American English vowels by experimentally inducing an alternation in speech style. They concluded that targets are not invariant attributes of vowels but adaptively specified so as to become sufficiently contrastive in the given communicative situation (see also ref. 5). It is the second purpose of this paper to examine a sample of spontaneously produced speech for evidence compatible with Lindblom and Moon's conclusion.

Method

We used for this experiment approximately one hour of recorded speech produced by a male native speaker of the Stockholm dialect of Swedish (subj. JS). The greater part of the recording is a lively monologue supported by brief questions and comments. The recording was carried out in a sound-shielded recording room (Revox PR/99 tape-recorder at 19 cm/s, Sennheiser 211/U microphone). For the present purpose, we selected the following phoneme sequences for analysis: /vi:s/ which in this speech sample occurs in the words vis 'manner', 'way', visum 'visa' and visa 'show'; and /ve:t/ which occurs in the verb form vet 'know(s)'. Both phoneme sequences consistently appear in primary-stressed position in this speech sample; lexical stress can thus be excluded at the outset as a potential source of spectral variation.

Broadband spectrograms were made of all instances of these sequences (Kay Elemetrics Digital Sona-Graph/7800, 300 Hz filter bandwidth, frequency range 0-5 kHz; Sona-Graph Printer/7900). The spectrograms were segmented at major acoustic discontinuities and measured for segment durations and formant frequencies. Onset and offset formant frequencies were measured at the respective first and last glottal pulse pertaining to the vocalic segment. Formant frequencies were measured at turning-points, if any, otherwise half way through the duration of the vocalic segment (see ref. 3 for details).

Results and discussion

The difference between maximum and offset second formant frequency, F_2(max-offset), as a function of duration of /e:/ in the sequence /ve:t/ is shown in Figure 1. The statistical correlation between the variables is $r=0.96$ (statistically significant at $p<0.01$). Further calculations show slightly lower but statistically significant spectral-durational correlations for F_2(max-onset) and F_2max. There is thus a clear effect of segment duration on the spectral parameters relating to F2, whereas the corresponding parameters relating to F3 are not highly predictable from duration.

Figure 2 illustrates the difference between maximum and onset third formant frequency, F_3(max-onset), as a function of the duration of /i:/ in the phoneme sequence /vi:s/. The statistical correlation between these variables is $r=0.74$ (significant at $p<0.01$). Further analysis reveals significant correlations also for F_3(max-offset) and F_3max as well as for the corresponding parameters related to the second formant. We thus find further evidence for an effect of segment duration on these spectral parameters.

However, both sets contain several data points whose variation along the spectral dimension is considerable within narrow time intervals. This is particularly evident for the set represented in Figure 2. Thus, a simple duration-dependent model does not account for the full range of spectral variation in these data.

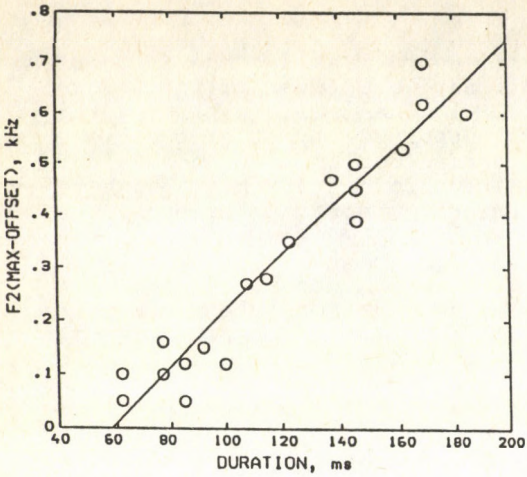


Figure 1. Difference between vowel F_2 maximum and offset (Hz) for /e:/ as a function of vowel duration (ms) in instances of the sequence /ve:t/.

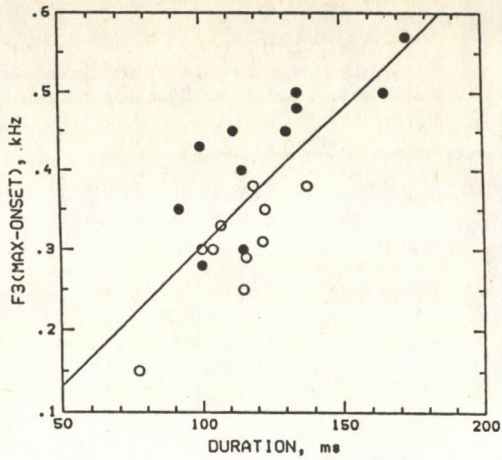


Figure 2. Difference between vowel F_3 maximum and onset (Hz) for /i:/ as a function of vowel duration (ms) in instances of the sequence /vi:s/. Unfilled circles: adverbial dummy phrase; filled circles: remaining phrases. Further explanation in text.

The spectral dispersion pattern in Figure 2 draws attention to the above-mentioned findings of Lindblom and Moon (6). They found durations and formant frequencies in vowels to vary independently when the subject alternated between "clear" and "citation form" speech. Their interpretation was that clear speech involves reorganization of phonetic gestures such that sufficient salience is given to contrastive signal information (cf. 1,2,5). Even though the data reported in the present paper come from an entirely different, spontaneously spoken sample, they nevertheless add some support to the general claim that an independent control dimension must be posited to account for varying degrees of phonetic elaboration and reduction in vowels, and that variation along this dimension may be understood in terms of communicative requirements. The evidence is the following:

The unfilled circles in Figure 2, which are mainly located below the regression line, represent instances of the sequence /vi:s/ occurring in the adverbial "dummy phrase" <u>på något vis</u> ('in a way', 'sort of'). Rather than adding new semantic information, such phrases seem to reflect the subject's own attitude to the current message. In this sense, they are semantically non-focal in the utterance context. It is conceivable that, in

the speaker's utterance plan, separate, duration-independent vowel targets are specified according to the semantic weight of the word or phrase in which the vowel in question is to appear. A greater semantic weight would then lead to a spectrally more elaborate vowel, and vice versa. At this point, however, this is rather speculative. Alternative approaches are possible, and much more data are needed to give conclusive support to any one of them. Nevertheless, the present findings are encouraging in that they clearly suggest that a fair amount of the phonetic systematicity observed in laboratory speech is readily discernible also in spontaneous speech. Thus, spontaneous speech can, and should, be subjected to further quantitative phonetic treatment. It is our conviction that future studies along these lines will add much valuable knowledge to our current picture of the speech production process.

Footnote

[1] "Speech Transforms" (Uttalstransformer), supported by The Swedish National Board for Technical Development (contract 88-02192P) and The Bank of Sweden Tercentenary Foundation (contract 86/109).

References

1. ENGSTRAND, O.: Articulatory coordination in VCV utterances - a means-end view. Reports from Uppsala University, Department of Linguistics (RUUL) 10. 1983.

2. ENGSTRAND, O.: Articulatory correlates of stress and speaking rate in Swedish VCV utterances. Journal of the Acoustical Society of America 83. 1988, 1863--75.

3. ENGSTRAND, O.--KRULL, D.: On the systematicity of phonetic variation in spontaneous speech. Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS) 8. 1988, 34--47.

4. LINDBLOM, B.: Spectrographic study of vowel reduction. Journal of the Acoustical Society of America 35. 1963, 1773--81.

5. LINDBLOM, B.--ENGSTRAND, O.: In what sense is speech quantal? Commentary on focus paper by K.N. Stevens, to appear in theme issue of Journal of Phonetics.

6. LINDBLOM, B.--MOON, S.-J.: Formant undershoot in clear and citation form speech. Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS) 8. 1988, 21--33.

7. NORD, L.: Acoustic studies of vowel reduction in Swedish. Speech Transmission Laboratory, Quarterly Progress and Status Report (STL-QPSR) 4/1986. Department of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm. 1986, 19--36.

SOME REMARKS ON UNDERSPECIFICATION

Bernhard HURCH

* Bergische Universität, Wuppertal *

0. The first edition of "Principien der Sprachgeschichte" goes back to 1880, the second edition is dated 1886. The 3rd chapter is dedicated to "Der lautwandel" and contains some important revisions from the second up to the 5th edition, on which the edition now available is based. One of the crucial problems for the neogrammarians and for the rigid conception of the **Lautgesetze** was the sporadic abrupt sound change. The role of this type of change differs considerably from one author to the other.[1] Whereas Brugmann (1886) argues for the inclusion of abrupt sound change in the **Lautgesetze**, saying that we cannot simply exclude what we are not able to explain at a certain moment, Baudouin rejects the neogrammarian concept of **Lautgesetze** altogether. H.Paul's position (from the second edition onwards) is somewhat more cautious: first he explicitly denies the parallelism between **Lautgesetze** and laws in the natural sciences and second he keeps abrupt sound change apart from the **Lautgesetze**.

1. It is obvious that exactly this type of phonological phenomenon had to represent a special challenge for nonlinear approaches.[2] One of the basic assumptions, formulated first in Goldsmith (1976) and taken over to all variants of nonlinear phonology, is that 'association lines do not cross'. As most of the abrupt sound change is of non-contiguous character (like metathesis and other types of permutation, distant assimilations, distant dissimilations, haplology, reduplications, migration processes etc.) the question arises how to explain within a nonlinear framework for example the change of initial *p- to a labio-

---

1 It is almost impossible here to review this dispute briefly. For the nearly complete lack of a theoretical discussion in the light of more recent phonological ideas - with the exception of Vennemann & Wilbur (1972) - I have to refer the reader to the originals particularly pertinent to the discussion at hand: Paul (1880 and 1886), Brugmann (1886), Wundt (1887), Jespersen (1884), Kruszewski (1884-86), Misteli (1888), Schuchardt (1885), Baudouin de Courtenay (1895).

2 I use "nonlinear" in the common sense as a cover term for the various recent contributions to phonological theory which try to overcome the classical vertical division of utterances into phonemes by positing different horizontally organized units, corresponding more or less to features, but which have a stronger autonomy than structuralist or SPE-phonology would admit. The term "nonlinear" is somewhat misleading as the organization of the single tiers is as linear as phonology has ever been (Hurch in prep.).

velar kw- resulting in Lat. quinque "five" (cfr. Gr. pénte, Osk.-Umbr. pumpe-, Lit. penki, Goth. fimf, O.Sl. pętĭ, Sanskr. pañca and so forth)[3]. This type of change, which Grammont (1933) treats under "dilation", shows an anticipatory effect of the labiality of the /p/ on the initial consonant, the labiality thus crossing the features of the intervening vowel and, in the example quoted, of the nasal consonant.

2. The recent answer to this crucial point is "underspecification". Since structuralism it has been discussed whether segments have to be specified with respect to all features they are alligned with or not. Archangeli (1988) gives, concerning this discussion within the generative framework, a rather satisfactory resumée. In addition, she introduces basically two concepts which on the one hand should present evidence for radical underspecification (versus contrastive specification) and on the other hand (at least the first one) should offer an explanation for the problem at hand: transparency and asymmetry. Transparency denotes the proposal that the adjacency of non-contiguous segments can be established by radical underspecification insofar as the intervening segments are not specified (and thus transparent) with respect to the feature involved in the substitution.

3. Archangeli (1988)[4] offers in her "empirical" part two examples for transparency, both of which seem to be somewhat 'inaccurate' - and the inaccuracy concerns crucial points of (radical) underspecification theory.

3.1. The Sanskrit retroflexion is more complex than presented by Archangeli,[5] as there are two more cases in which the retroflexion does not take place. These are if a) a nasal /n/ in the position of being retroflexed is followed by a stop (e.g., granthí- "node"), and b) if the retroflex consonant which spreads this property is a stop (either nasal or oral) (e.g., kaṭhina- "steep", maṇinā "perl").[6] These two aspects ommitted by Archangeli are interesting insofar as they demonstrate that the retroflexion is blocked exactly where we find phonetic reasons for it: The spreading of a feature is not in itself something that happens but it depends on the nature of the feature and it depends

---

3 There are analogous velar forms in some Celtic languages, cfr. Ir. cóic, but Corn. pymp, Br. pemp, etc.

4 I consider it admissible to concentrate particularly on Archangeli (1988) as her study is the introduction of the guest editor to a special issue of the Phonology Yearbook 5.2.

5 The main source for Sanskrit - I conclude this from the presentation of the data - seems to have been Schein & Steriade (1986), as for the following Latin example expressively only Steriade (1987). I don't think that Archangeli wants to limit the Sanskrit example to deverbalized nouns formed with -ana, when the phenomenon she is describing is much more general.

6 In the following, for reasons of space, I will be able only to discuss point b) in some detail.

on the particular configuration/interaction of features within a segment. It further shows that the tiers[7] are not so autonomous as retroflexion can spread from a segment which is continuant but it cannot spread from a segment which is non-continuant. The reason is simply that the nature of a stop includes the oral release of the stop position and there is no sense in spreading a position which articulatorily has already been given up. I am perfectly aware that these considerations are not primarily arguments against underspecification but they should offer evidence for a more phonetically based conception of features (a point which ultimately is expressedly excluded by Clements [1985]) and it should show that radical underspecification is problematic, as it might be true that segments are identifyable even if radically underspecified but that a specific grammar (including substitutions) might require a higher degree of specification.[8]

3.2. The problem concerning the second example Archangeli (1988) gives under the heading 'transparency' is of a very different nature. She discusses the distribution of the Latin suffixes -alis and -aris, the use of which being dependent on whether the last liquid in the stem is /r/ or /l/ respectivly. This suffix seems not to have been too productive in Latin and, no wonder, most of the examples Archangeli quotes directly from Steriade (1987:351) are lexicalized forms. It is difficult to go into detail as the argumentation line is not completely clear. It is no infrequent case that the choice of allomorphs depends on morphonological variables. But it is equally not infrequent that these criteria which operate in the choice of alternative suffixes are not limitations over lexical representations but possess validity in one particular morphological circumstance. And this is exactly the case in the Latin example. We clearly have forms such as lilium "lily" which do not undergo any kind of dissimilation. And furthermore doublets arise like litteraris and litteralis. The point why the discussion of the exclusion of many of the cases of non-contiguous abrupt and all cases of sporadic sound change arose among the neogrammarians is exactly the same that makes the -aris/-alis discussion for Latin phonology (and thus for underspecification) obsolete: the distinction between phonology and morphonology is not a decision open to grammatical theory but is a distinction imposed by grammar.

4. The difference between **segment filling** and **segment changing** rules corresponds roughly to the difference between paradigmatic and syntagmatic processes in natural phonology. And it is the limitation of the former which makes up, under typological

---

7 I am not completely sure whether 'retroflex' has to be situated under the manner node (insofar as it exists - Clements [1985]) or under the place node. There are arguments for both solutions.

8 And this is exactly the point natural phonology has always been arguing for: the choice of features (and thus in the present discussion: their specification) must depend on the processes operating in a given phonology.

and language specific conditions, a given system. But one characteristic of paradigmatic processes is that they are present at any step of the derivation, conflicting or not. Latin /n/ is not-labial or coronal or dental or alveolar or whatever articulatory description it might bear, at any step of the derivation and not only in the contrastive sense but because /n/ undergoes a series of assimilations (point of articulation, nasalizations etc.) which neither the labial nor the palatal nor the velar do. And even if we force a markedness relation to be expressed in terms of (a lack of) feature specifications, Latin /n/ has to bear a place feature, intervening in the above cited anticipatory velarization. Markedness, default values and the like are results of processes, as Stampe (1973) has shown, they cannot bear an explanatory role themselves.

*****
***
*

REFERENCES:

Archangeli, D. (1988) Aspects of underspecification theory. PHONOLOGY YEARBOOK 5:183-207

Baudouin de Courtenay, J. (1895) VERSUCH EINER THEORIE PHONETISCHER ALTERNATIONEN. Straßburg

Brugmann, K. (1886) VERGLEICHENDE LAUT-, STAMMBILDUNGS- UND FLEXIONSLEHRE. vol.I of Brugmann & Delbrück, GRUNDRISS DER VERGLEICHENDEN GRAMMATIK DER INDOGERMANISCHEN SPRACHEN. Straßburg

Clements, G.N. (1985) The geometry of phonological features. PHONOLOGY YEARBOOK 2:225-252

Goldsmith, J. (1976) AUTOSEGMENTAL PHONOLOGY. IULC, Bloomington

Grammont, M. (1933) TRAITE DE PHONETIQUE. Paris

Hurch, B. (in prep.) Dictionnaire des idées linguistiques reçues. ms.

Jespersen (1884) Zur Lautgesetzfrage. INT. ZEITSCHR. FUR ALLG. SPRACHW. III:188-216

Kruszewski, N. (1884-86) Prinzipien der Sprachentwicklung. INT. ZEITSCHR. FUR ALLG. SPRACHW. I and III:295-307, 145-187

Misteli (1888) Lautgesetz und Analogie. ZS. FUR VOLKERPSYCH. UND SPRACHW. XI.4:365-475

Paul, H. (1880, 1886[2]) PRINCIPIEN DER SPRACHGESCHICHTE. Halle

Schein, B. & D. Steriade (1986) On geminates. LI 17:691-744

Schuchardt (1885) UBER DIE LAUTGESETZE: GEGEN DIE JUNGGRAMMATIKER. Berlin

Stampe, D. (1973) On chapter nine. in M.Kenstowicz & Ch. Kisseberth (eds.) ISSUES IN PHONOLOGICAL THEORY, Den Haag, 44-52

Steriade, D. (1987) Redundant Values. CLS 23/2:339-362

Vennemann, Th. & T. Wilbur (1972) SCHUCHARDT, THE NEOGRAMMARIANS, AND THE TRANSFORMATIONAL THEORY OF PHONOLOGICAL CHANGE, Frankfurt

Wundt, W. (1887) Über den Begriff des Gesetzes, mit Rücksicht auf die Frage der Ausnahmslosigkeit der Lautgesetze. PHILOSOPHISCHE STUDIEN 3:195-215

ON VOWEL LENGTH VARIABILITY IN HUNGARIAN

Ilona KASSAI
Linguistics Institute, Hungarian Academy of Sciences, Budapest, Hungary

Introduction

Extensive variation in the phonetic realization of short and long vowel quantities has been, from the start, one of the main concerns in descriptions of the sound shape of Hungarian. The high long vowels /i:/, /y:/, /u:/ are especially affected by a shortening tendency while mid vowels /ø/ and /o/ are subject to both shortening and lengthening. Since the low vowel pairs ɛ/e: and ɔ/a: show qualitative differences as well, their durational variation is not as noticeable as in the case of mid and high vowels. Though among the approaches "therapeutic" ones have had a predominance as opposed to the "diagnostic" approach, this latter has yielded remarkable results of both auditory and instrumental analyses introducing sociological factors in the selection of the informants (1, 2, 3). The current paper is a continuation of this line of research in that it undertakes to analyse the above phenomenon as a function of speech tempo, speech style, written form and phonological awareness.

Materials and method

The materials of 10 secondary school teachers of over 50 years and 10 vocational trainees around age 16 were selected for auditory analysis from the quota sample of Version Two of the Budapest Sociolinguistic Interview, a preliminary corpus for the purposes of the Survey of Spoken Hungarian. I chose these two groups of informants from among a total of five socio-economic groups with 10 persons in each as they represent extreme values within the sample. They differ in age, education, cultural level, social position and even in their relation to language in that teachers' vocation is to mediate the norms of standard Hungarian towards their pupils while vocational trainees, in their turn, are expected to accept those norms. Consequently, if the analysis explores convergencies in their speech performance, these convergencies are to be considered as language specific rather than speaker specific. The 2-hour spoken corpus of each interview has been collected by a number of different methods: reading out minimal pairs, word lists and short passages, the latter both at normal and fast tempo, data elicitation by questions, listening tests, fill-in tasks and guided conversation. The test-like part of the interview makes direct comparisons across informants possible, while the other part, constituted by continuous speech, may support additional evidence for or against the phenomena studied systematically in the first part.

With respect to vowel length the first part of the interview makes it possible to examine the following questions:

(1) what is the impact of speech tempo on vowel length variability?
(2) how does typewritten text influence the speakers reading performance?
(3) how do speakers achieve the quantity contrasts in the various speech styles ranging from formal to casual speech according to the decreasing amount of self-monitoring?
(4) to what extent are speakers aware of vowel length variability?

The questions will be answered on the basis of the data processed from the

auditory evaluation of the actual realizations of short and long quantities.

Results

Effect of tempo. - The interrelations of vowel length and speech tempo could have been analysed in 7 passages read at normal and fast rate by the informants of the two socio-economic groups. As, however, most speakers could not speed up their reading tempo, I selected two successful readers from each group. The data relative to their performance are shown in % in Table 1.

Table 1. 7104=code number of the informant; +=long vowel with long duration; -=long vowel with short duration; N=normal tempo; F=fast tempo

Duration	Teachers				Vocational trainees			
	7104		7412		7514		7515	
	N	F	N	F	N	F	N	F
+	66	44	59	43	33	31	24	14
-	34	56	41	57	67	69	76	86

As becomes obvious from the table, the shortening effect of speech tempo is more marked among the teachers than among the vocational trainees. Nevertheless, this conclusion has to be followed by another one, namely that vocational trainees, at normal rate, pronounce nearly twice as many short wowels instead of the long ones than do teachers. Therefore, the right conclusion has to be formulated so that tempo has a greater effect on the realization of long quantity in the performance of the teachers simply because in vocational trainees' fast reading there is hardly any long vowel left for fast rate to shorten.

Effect of written form. - Until about six years ago the keyboards of Hungarian typewriters lacked three keys: í, ű, ú. Instead of these high long vowels only their short equivalents (i, ü, u) could be typed. It has been claimed several times that this deficiency of the keyboard has an influence on people's speech i.e. makes them use short vowels intead of standard long vowels, thus accelerating the tendency to shorten these vowels. Therefore we can examine whether the informants will read out words differently if they are spelt on typewriters with the old and the new Hungarian standard keyboard (e.g. hosszu - hosszú 'long') or if they are spelt according to the previous or the recent edition of the Orthographical Rules of the Hungarian Academy (e.g. zsüri - zsűri 'jury'). The data of the analysis are contained in Table 2. According to the data both in teachers' and vocational trainees' reading performance one can see hardly any difference attributable to differences in spelling. Among the teachers in 3 spelling pairs the actual realization contradicts the norm and only 2 pairs conform to the canonical form. Among the vocational trainees the picture is the same: the 4 cases not obeying the written forms show 3 substandard and 1 standard solutions.

Table 2.

Spelling pairs		Teachers		Vocational trainees	
		short	long	short	long
irjanak		3	7	6	4
írjanak	'they write'	3	7	2	8
kinlódjanak		2	8	7	3
kínlódjanak	'they suffer'	-	10	2	8
zsüri		7	3	2	8
zsűri	'jury'	3	7	1	9
hasonszőrü		5	5	8	2
hasonszőrű	'similar'	6	4	10	-
ujabban		4	6	6	4
újabban	'recently'	7	3	6	4
gyanu		6	4	10	-
gyanú	'suspicion'	3	7	6	4
hosszu		4	6	9	1
hosszú	'lengthy'	7	3	9	1

Effect of speech styles. - Looking carefully at the values one can state that the performance of both groups is fairly even along the different tasks despite of the decreasing amount of self-monitoring they require. To put it differently, the proportion of shortened vowels seems to be held constant along the formal--casual axis. If, however, we compare the two groups it appears that the teachers perform in a more varied way than do vocational trainees which means that they adapt themselves to the peculiarities of the different styles to some extent while vocational trainees do not.

Linguistic awareness and vowel quantity. - In the sociolinguistic interview three listening tests consisting of word pairs and administered through headphones serve to explore the relation between the informants' linguistic awareness and language use. In the 'Same or different?' test the informants have to decide if the words in a pair mean the same or not. There are examples for different meanings and identical meanings with variant pronunciations. The answers show if the subjects are aware of the phonological function of vowel quantity. The data make it clear that the meaning differentiating function of quantity in high vowels is more or less obscure for the informants: more for vocational trainees and less for teachers.

In the 'Which is correct?' test out of two different pronunciations of certain words the correct pronunciation has to be identified. The third test called 'How do YOU usually say it?' asks for the forms used by the informant. These two tests jointly yield the linguistic uncertainty index of the informants since it may happen that the answers to the two questions are in conflict with each other. Indeed, the data illustrate a considerable degree of uncertainty as regards of the knowledge of quantity in the different words. The correctness judgements of the vocational trainees seem to be based on their pronunciation practice: if they pronunce a word consistently with long vowel, e.g. hívő 'believer', they judge this form correct. If, on the contrary, they usually say a word with a short vowel, e.g. bölcsöde, this form is taken to be correct. The teachers seem to hesitate, they are less "straightforward" in their decisions. The tendencies emerging from the data are further strengthened by spontaneous comments accompanying listening tasks. Let us quote a few of them. "I had trouble with long í, ú, ű."; "Bölcsöde 'day care center'. Bölcsőde (self-correction). Sorry. I usually say bölcsöde, with short ö.

Unfortunately."; "I tend to pronounce kőrut 'boulevard' [for körút] myself, since all Budapest says it like this. I call it the Budapest dialect." These remarks are precious for they point to the fact that contradictory judgements may be conscious. And, as they have been made by teachers only, one can deduce that the way vocational trainees use their language is far less conscious.

As a final step, I have arranged the words gathered from the 7 passages read at normal tempo according to the frequency of shortening of their long vowel in the two groups (lengthening has been represented by a single item). The two word lists confirm the tendency stated with respect to tempo, namely that the vocational trainees shorten the long vowels considerably more often than the teachers: there are 18 words out of 56 in which all vocational trainees pronounce a short vowel instead of the long one; in the list of the teachers there are only 2 such words. If, however, we disregard numerical proportions we can state that the rank order of the different words is the same in both lists. With a closer analysis of this rank order one can read out those tendencies that have been formulated on the issue, e.g. prevocalic position favors shortening more than the other positions, in unstressed syllables of polysyllabic words shortening is more marked than in the stressed syllable, stressed long vowels are more stable in monosyllables than in polysyllabic words etc. It also turns out that on the basis of the rules mentioned one cannot predict the actual duration of the two contrastive quantities, as all conditions being equal there is a great variability in terms of individual words.

Discussion

According to the data, the variables involved in the analysis have greater influence on the performance of the teachers than of the vocational trainees. Moreover, the former group of informants is aware of the uncertain use of contrastive quantity while the latter group is not. Nevertheless, the tendency to shorten long vowels and to lengthen short ones appears to be more marked in the speech production of the vocational trainees, i.e. they more consistently deviate from the standard.

As the variability of short and long vowels seems to escape phonetic rules we have to look for some other explanation. The metric system of Hungarian is likely to be a promising point of departure as it is able to harmonize durational and accentual metrics in order to describe the rhythm of the language. E.g. there seems to be a requirement for the syllable not to be long both by its position and its vowel. If vocalically long syllables become long by positon too, the long vowel has a strong tendency to shorten. A tentative analysis of a few words of varying length appears to reveal that,as long as morpheme structure constraints are not thereby violated, the word as a unit tends to be constructed of full metric feet and of feet which are accepted in Hungarian poetry. This hypothesis can readily be verified through the analysis of the continuous speech sample of the sociolinguistic interview.

References

1. Fónagy, I.: Über den Verlauf des Lautwandels. AlinguH 6. 1956, 173--248.
2. Magdics, K.: Kürzung der unbetonten Vokale in der ungarischen Umgangsprache ZPhon 14. 1961, 21--44.
3. G. Varga Gy.: Alakváltozatok a budapesti köznyelvben. Budapest, 1968.

THE PROTO-SEMITIC SIBILANTS

Marina MEPARISHVILI
Tsereteli Institute of Oriental Studies
Georgian Academy of Sciences, Tbilisi, USSR

During the last period in Semitic comparative linguistics the increase of interest in the studies concerning the reconstruction of phonological systems of parent-language as well as of its different branches is observed. Attention is also paid to their relationship from the viewpoint of diachronic linguistics.

The comparative investigation of the vocabulary of modern spoken Semitic languages (have been up till now only slightly studied from the comparative point of view), though the data of these languages may be used in elaboration of particular principles for the comparative phonology. They works of remarkable orientalists W. Leslau (such as dictionaries of Harsusi, Jibbali/Shari) should be especially noted. The data of the dictionaries of the languages noted above can be successfully used in the research of the concrete problems of comparative Semitic phonology.

It should be admitted, that reflexes of some Proto-Semitic phonemes are fluctuated in different Semitic languages. It is necessary to define them more precisely; first of all the sibilants should be regarded as same phonemes. We'll consider the correspondences for Semitic sibilants in different languages.

There are used the following abbreviations for the languages: Sem-Semitic, Arab-Arabic, Eth-Ethiopic, (Ge)-Geez, (Te)-Tigre, /Tna)-Tigrinya, (Amh)-Amharic, (Har)-Harari, (Gaf)-Gafat, (Gur)-Gurage, SA-South-Arabian, (Ep)-Epigrafic, (Hrs)-Harsusi, (Mh)-Mehri, (Sh)-Shahri/Jibbali, (Soq)-Soqotri, Hbr-Hebrew, Ug-Ugaritic, Akk-Akkadian, Aram-Aramaic.

In the series of interdental sibilants /ḏ/, the voiced member of the phonological opposition, preserves its original phonetic significance in Arabic and South-Arabian languages (with the exception of Soqotri, where voiceless dental ḏ corresponds to Sem /ḏ/, the voiced member of the phonological opposition, preserves its original phonetic significance in Arabic and South-Arabian languages (with the exception of Soqotri, where voiceless dental ḏ corresponds to Sem /ḏ/). In Hebrew, Akkadian and Ethiopic There occurs /z/ and in Aramaic languages - /d/.

Sem* *ʾuḏn "ear"

Arab *ʾuḏn "ear" Eth(Ge) ʾezeⁿ "ear" [6], Eth(Te) ʾezeⁿ [16], Eth(Tna) ʾezni, Eth(Gaf) ʾezn [13], Eth(Har) ʾuzuni [12], SA(Ep) ʾḏn "ear;listening; obedience" [3], SA(Hrs) he-yḏen "ear" [7], SA(Mh) he-yḏen, SA(Sh) ʾiḏen [8], SA(Soq) ʾedhan [11], Hbr ʾozen, [10], Akk uznu [17], Aram ʾdn [5], Ug ʾdn [1].

Common Semitic /ṯ/ is the voiceless correlate of the voiced interdental sibilant /ḏ/. This phoneme prezerves its original phonetic significance in Arabic, South-Arabian and Ugaritic. In Hebrew, Akkadian is /š/, in Ethiopic - /s/, in Aramaic - /t/.

Sem* ṯbr "to break"

Arab ṯabara "to do harm" [4], Eth(Ge) sabara "to break, crush", Eth(Te) säbrä "to break", Eth(Tna) säbra, Eth(Amh) säbbärä [2], Eth(Har) säbära, Eth(Gur) säbärä "to break, destroy" [14], SA(Ep) ṯbr "to damage s.o., to harm, destroy; to defeat an enemy", SA(Hrs) ṯebōr "to break in pieces",

SA(Mh) tebor,SA(Sh) Tōr,Hbr šebar "to break, break in pieces", Ug tbr "to break", Aram tebar, Akk šabburu "broken".

The third, glottalized or "emphatic" member of the considered opposition is /ṭ/. The correspondences for this common Semitic phoneme are: /ẓ/ in Arabic, SA(Ep) and Ugaritic, /ḏ/ in South-Arabian (/ṭ/ in Soqotri), /ṭ/ in Aramaic. In Hebrew, Akkadian and Ethiopic occurs /ṣ/. But /ṣ/ was changed within the scope of the Ethiopian branch: in South-Ethiopic there are three correspondences: ṣ,glottalized dental /ṭ/ and glottalized voiceless affricate/č̣/.

Sem* tfr "nail,claw"
Arab 'uẓfur "nail",Eth(Ge) ṣ^{e}fr "fingernail",Eth(Te) ṣ^{e}fr , Eth(Tna) ṣ^{e}fri,Eth(Amh) ṭefer "nail",Eth(Har) ṭ^{e}fre,Eth(Gaf) ṣ^{e}fra,Eth(Gur) ṭ^{e}f^{e}r, SA(Ḥrs) ḏefir,SA(Mh) ḏēfīr,SA(Sh) ḏefer,SA(Soq) ṭefer,Hbr ṣippor-en "nail (finger,toe)",Aram ṭ^{e}pra "nail,claw",Akk ṣupuru.

In the series of Semitic simple sibilants /z/ is the voiced member of the phonological opposition, which has the only correspondences in various Semitic languages, namely /z/.

The voiceless correlate of Sem /z/ in the series of the simple sibilants is /s/. In every Semitic languages it appears as /s/.

The South-Semitic /s/ reflects two Proto-Semitic phonemes -/s/ and /š/, as a result of dephonologization of PSem /š/ in South-Semitic languages. In the sec ond case there is s$_{4}$ in SA(Ep), /š/ - in Akkadian, Hebrew, Ugaritic, Aramaic and /s/ in South-Semitic (Arab., Eth.,SA).

Sem* npš "soul, spirit, breath"
Arab nafs "soul,spirit",Eth(Ge) näfäsä "to breath, blow", Eth(Te) näfäsä "to breath, to have a soul", Eth(Amh) näffäsä "to breath",Eth(Gaf) n^{e}f^{e}s "wind",Eth(Gur) näfs "soul",SA(Ep) nfs, ,SA(Hrs) nefes-et,SA(Sh) nefs-et,Hbr nepeš "breath, soul, personality", Aram napšā "soul".

/Ṣ/ is the glottalized (or "emphatic") member in the series of Semitic simple sibilants. The principal correspondence for this phoneme in different languages is /ṣ/. It is a voiceless glottalized sibilant, which shows definite stability and does not undergo any change. The Ethiopic languages are exceptions in which /ṣ/ somewhat modified. In North-Ethiopic (Geez, Tigre,Tigrinya) /ṣ/ is preserved, while in South-Ethiopic there are: /ṣ/, /ṭ/ and /č̣/.

Sem* ṣwm "to fast"
Arab ṣāma "to fast"Eth(Ge) ṣama "to fast,abstain from food", Eth(Te) ṣomä "to fast", Eth(Amh) ṭomä "fast", Eth(Gaf) ṣima "to fast",SA(Ḥrs) ṣōm,SA(Mh) ṣōm,SA(Sh) ṣum,SA(Soq) ṣiom,Hbr ṣwm "to fast" Aram ṣam.

In the series of lateral sibilants the voiced member of a phonological opposition is absent, while the voiceless and the glottalized ones are present. The common Semitic voiceless lateral sibilant /ŝ/ in Ethiopic and Arabic was replaced (probably owing to a difficulty of its pronunciation) by the voiceless pre-palatal sounds, the phonetic character of which is simpler a that of the original sound. By that time /š/ was "released" from the system of sibilants of these languagesas a consequence of change of Sem /š/ into /s/ in South-Semitic branch. In modern South-Arabian the correspondence for the considered common Semitic phoneme is a voiceless lateral sound transcribed by scholars as ś,in SA(Ep) is s$_{2}$, in Hebrew - /ś/, in Aramaic - /s/, in Akkadian and Ugaritic - /š/.

Sem* ŝyb "to be old, have grey hair"

Arab šayb "grey hair", Eth(Ge) šeba "to have grey hair", Eth(Te) šayb "old, grey", Eth(Amh) šib-at "grey hair",Eth(Har) šib-at, Eth(Gur) $š^{e}$b-et, SA(Ep) $s_2$yb "to be old, grey-headed", SA(Hrs) śayb "white hair", SA(Mh) śayb, SA(Sh) eśśeb "to have grey hair", SA(Soq) śaybi "old man", Hbr śāyb "grey-headness", Ug šb "old man", šb-t "grey-headness", Akk šibu "old man", Aram sābā "old".

/$\hat{ṣ}$/ is the voiceless glottalized correlate of common Semitic /$\hat{s}$/. The considered phoneme is the most modified phoneme is the system of Proto-Semitic sibilants. This phoneme doesn't occur in any Semitic language in its original phonetic significance. The Sem /$\hat{ṣ}$/ in Arabic corresponds to an emphatic dental /ḍ/, in SA languages -- the voiced lateral sibilant transcribed by T.M. Johnstone as ź, in Hebrew, Ugaritic and Akkadian -- /s/,' in Aramaic. In Ethiopic Sem /$\hat{ṣ}$/ merged with an Ethiopian /ṣ/ < Sem /ṣ/, in South-Ethiopian languages it was modified as well as Sem /ṣ/.

Sem* $\hat{ṣ}$bṭ "to hold, seize"

Arab ḍabaṭa "to hold, snatch, seize", Eth(Ge) ḍabaṭa "to hold fast seize, catch", Eth(Te) ṣäbṭä, Eth(Tna) ṣäbbiṭä "to seize, catch", Eth(Amh) č̣äbbäṭä "to hold, seize, catch", SA(Hrs) źeybeṭ "to take", SA(Mh) źāṭ, SA(Sh) źōṭ, SA(Soq) źeybeṭ "debtor, whose property is under arrest", Hbr ṣebat "to hold out", Ug ṣbt "hold".

The analysis of the phonetic significance of considered phonemes as well as of the new data of experimental phonetics and comparative Semitic phonology [7, 9, 15, 18, 19] allows to reconstruct nine original sibilant phonemes, which could be united in a system.

The system of Proto-Semitic sibilants is reconstructed in the following way:

	voiced	voiceless	glottalized
interdental	$\underline{d}$	$\underline{t}$	$\underline{\dot{t}}$
simple (alveolar)	z	s	ṣ
(prepalatal)		š	
lateral	-	$\hat{s}$	$\hat{ṣ}$

The Proto - Semitic paradigmatic system of sibilants is represented by three-member oppositions (voiced : voiceless : glottalized) in horizontal axis. In its turn the opposition in vertical axis is composed too. Particularly should be noted a complicated relations between the members of the "redundent" four-member opposition in the series of voiceless sibilants. In the majority of Semitic languages this opposition turned into the three-member opposition, moreover, in Ethiopic and Akkadian it turned into the two-member opposition s:s. However, these transformations result in different reasons. In Akkadian Proto-Semitic oppositions š:t and š:$\hat{s}$ were neutralized (t > š, $\hat{s}$ > š). Thus, Proto - Semitic opposition $\underline{t}$:s:š:$\hat{s}$ was changed into s:š. As to Ethiopic, the Proto - Semitic opposition s:š (Sem š > Eth s) was at first neutralized, and then Proto - Semitic phonemes $\hat{s}$ and $\underline{t}$ were dephonologized. Thus, according to these modifications, the Proto - Semitic four-member opposition $\underline{t}$:s:š:$\hat{s}$ was changed in Ethiopic into the two-member opposition s:š.

The reconstructed system could be considered from the viewpoint of the problem of markedness. The series of simple sibilants demonstrates stability, while the lateral and interdental ones are modified and merged with simple sibilants. Thus, this series should be regarded as unmarked. On the other hand, the series of interdental sibilants is marked in relation to the series of simple sibilants, but unmarked in relation to the series of lateral sibilants. The inference in based on the fact of absence of the voiced lateral. Moreover, the other two are preserved only in South-Arabian, while interdentals are preserved in South-Arabian, Arabic and Ugaritic.

While comparing various branches of Semitic languages with the Proto-Semitic, it should be noted, that the system of sibilants of South-Arabian branch is the most preserved in the whole language-family.

REFERENCES

1. AISLEITNER, A.: Dictionaire der ugaritischen Sprache, Berlin, 1963
2. BAETEMANN, J.: Dictionaire amarigna-français, Dire-Daqua, 1929
3. BEESTON-GHUL-MÜLLER,RYCKMANS: Sabaic Dictionary, Loovan-la-Neuve, 1982
4. BIBERSTEIN KAZIMIERSKI, A.: Dictionnaire arabe-francais, Paris, 1860
5. BROCKELMANN, C.: Lexicon Syriacum, Halle, 1928
5. DILLMANN, A.: Lexicon linguas aethiopicae, Lipsiae, 1865
7. JOHNSTONE, T.M.: Harsusi Lexicon and English-Harsusi Index, London, Oxford University Press, 1977
8. JOHNSTONE, T.M.:Jibbali Lexicon, London, Oxford University Press, 1981
9. JOHNSTONE, T.M.: Contrasting Articulation in Modern South Arabian languages, Hamito-Semitica, The Hague-Paris, 1975, 155--159
10. KOEHLER, L.- BAUMGARTNER, W.A.: Dictionary of the Hebrew Old Testament in English and German, Leiden, 1958
11. LESLAU, W.: Lexique soqotri (sudarabique modern), Paris, 1938
12. LESLAU, W.: Etymological Dictinary of Harari, University of California Press, Berkeley and Los Angeles, 1963
13. LESLAU, W.: Etude descriptive et comparative du Gafat, Paris, 1956
14. LESLAU, W.: Etymological Dictionary of Guzrage (Ethiopic), vol.I. Wiesbaden, 1979
15. LESLAU, W.: What is a Semitic Ethiopian Languages? Hamito-Semitica, The Hague-Paris, 1975, 129-131
16. LITTMANN, E. --HÖFFNER, M.: Wörterbuch der Tigre-Sprache, Wiesbaden, 1962
17. SODEN, W. von: Akkadischen Handwörterbuch, Wiesbaden, 1965-1974
18. SWIEGERS, P.A.: Phonological Analysis of the Harsusi Consonants, Arabica, XXVIII, 2--3, 1981, 358--361.
19. SWIEGERS, P.A.: Note on the Phonology of Old Akkadian, Orientalia Lovaniensia Periodica, 11. Leiven, 1980, 5--9.

THE EXACT DOMAIN OF CONSONANT DEGEMINATION IN HUNGARIAN

Ádám NÁDASDY
Department of English
Eötvös University, Budapest, Hungary

In Hungarian practically all consonants may occur short (single) or long (geminate). Geminates are in phonemic opposition with short consonants (as is normally shown by spelling):

hall [hɔll] 'he hears' ↔ hal [hɔl] 'fish'
kassza [kɔssɔ] 'cash desk' ↔ kasza [kɔsɔ] 'scythe'

There are various restrictions on the occurrence of geminates. Trivially, they are excluded from the word-initial position. More interestingly, available descriptions of Hungarian - very-traditional (Papp 1966), classical structuralist (Hall 1944) and generative (Vago 1980) alike - contend that geminates may not occur next to another consonant. They say or at least imply that this restriction applies everywhere within the phonological phrase regardless of boundaries. This makes it look like a late postlexical rule; my aim is to refine and limit the power of this rule of Degemination (Deg).

Certainly, geminates within the morpheme (=underlying geminates) may occur only in intervocalic (kassza) or final-postvocalic (hall) position. The latter are regularly degeminated (shortened) when the next formative begins with a consonant (this Deg is not shown in spelling). Eg:

hallva [hɔlvɔ] 'hearing'
hall minket [hɔlminkɛt] 'hears us'

It is, however, an overgeneralization to say that geminates may never occur next to another consonant. I will show that configurations of $C_1C_1C_2$ or $C_1C_2C_2$ do occur in Hungarian in certain well-definable environments. Let us first survey the logically possible surface combinations of a geminate being flanked by another consonant. Since this may not happen within the morpheme, there will always be some kind of boundary within such a sequence.

	left-flanked geminate	right-flanked geminate
• undivided geminate	(1) *X-CC (várt)	(2) *CC-X (hallva)
• divided geminate	(3) XC-C (talppont)	(4) C-CX (széppróza)

(CC=geminate; X=flanking consonant; - = boundary)

Types (1) and (2) symbolize geminates not divided by a boundary: these are underlying (intramorphemic), and not the result of any process. Types (3) and (4) symbolize derived geminates whose two "halves" are divided by a boundary: these are the result of some concatenation process.

Underlying Geminates

These are obligatorily degeminated next to any consonant. Type (1) is vacuous, unless we posit the past tense suffix to be underlyingly /tt/ (Vago 1980), cf. ugrott 'he jumped', which would degeminate after a consonant: várt 'he waited' (< vár+tt).

Type (2) always regularly degeminates (hallva, hall minket) in all possible domains, both lexically and postlexically. Writers on Hungarian phonology have apparently had this classic type in mind when making their generalization.

Derived Geminates

When the geminate is created by the concatenation of two identical consonants, Deg is conditioned by morphological or phonological factors.

In Type (3) we have words ending in two different consonants (XC]) to which an identical consonant is attached. Here the nature of the boundary, ie. the level at which the gemination occurred, is relevant in determining whether Deg will take place or not.

(3a) Deg of XC-C is obligatory when the geminate is formed at some early lexical stratum through the addition of particular suffixes, such as: (3ai) /v/-assimilating suffixes; (3aii) "intimate" inflexions, ie. verb conjugations or noun possessives:

(3ai)	vers+vel	→	verssel	[vɛršɛl]	'with a poem'
	pont+vá	→	ponttá	[ponta:]	'into a point'
	akt+val	→	akttal	[ɔktɔl]	'with a nude'
(3aii)	küld+te	→	küldte	[kültɛ]	'he sent it'
	rajz+jon	→	rajzzon	[rɔjzon]	'it should swarm'
	kard+ja	→	kardja	[kɔrɟɔ]	'his sword'

That these geminates - before regularly degeminating - are formed at the lexical level is shown by /v/-assimilation, which does not apply to all /v/-initial suffixes, cf. ront-ván [rontva:n] 'destroying'. It is also remarkable that these early rules produce geminate affricates (ie. long affricates, with the stop phase lengthened), eg.:

	Bécs+vel	→	Béccsel	[be:ččɛl]	'with Vienna'
	fűt+jön	→	fűtsön	[fü:ččön]	'he should heat'

such geminate affricates being regular input to Deg when preceded by a consonant, eg.:

	korcs+val	→	korccsal	[korčɔl]	'with a mongrel'
	ront+jon	→	rontson	[rončon]	'he should destroy'

(3b) When the geminate is created by the addition of other elements, Deg of XC-C depends on the flanking consonant (the X).

(3bi) Deg is obligatory if X is an obstruent:

koszttól	[kostol] 'from food'
direkttermő	[-ɛktɛ-] 'a type of vine'

(3bii) No Deg occurs if X is a liquid or a glide:

talppont	[tɔlppont]	'foot-end'
szerb bor	[sɛrbbor]	'Serbian wine'

sztrájkkor	[strajkkor]	'during the strike'
sért talán	[še:rttɔla:n]	'offends perhaps'

The above is not invalidated by the fact that some frequent compounds are usually pronounced with degemination after a liquid or glide; these are instances of lexicalization, ie. are treated like cases of (3a) above. They are (irregularly) handled at an earlier stratum than where they belong grammatically:

párttitkár	[pa:rtitka:r]	'party secretary'
sporttárs	[šport(t)a:rš]	'fellow sportsman'
Holt-tenger	[holt(t)ɛngɛr]	'Dead Sea'

When X is a nasal, Deg is optional:

tankként	[tɔnk(k)e:nt]	'like a tank'

This may reflect the ambiguous nature of nasals: as non-continuants they side with obstruents, while as sonorants they side with liquids and glides.

Deg as in (3b) is a truly postlexical application of the rule, which operates across all boundaries. This is supported by the fact that it cannot apply to affricates: two adjacent affricates are not merged into a geminate postlexically and thus form no input to the rule. Compare the above forms Béccsel (long affricate resulting from morpho-phonological rule), korccsal (the same, with Deg according to (3a)), but

kulcscsináló	[kulč-čina:lo:]	'key maker'
narancs-juice	[nɔrɔnǰ-ǰu:z]	'orange juice'

which retain the two affricates, separately released.

In Type (4) we have words that begin with two different consonants ([CX), preceded by a word (or prefix) ending in the same C. Here the working of Deg is the same as in (3b) - that is, as long as the other conditions are met, it is immaterial which side of the geminate the flanking consonant stands on. Deg is not direction-sensitive - it just so happens that the "intimate" affixes involved in Deg (3a) are all suffixes, ie. they produce XC-C. Let us see some examples for Type (4):

(4i) Deg of C-CX is obligatory if X is an obstruent (this practically means words in /š/ or /s/ plus /p, t, k/):

kisstílű	[kišti:lü]	'petty'
olasz sztárok	[olɔsta:rok]	'Italian stars'

(4ii) No Deg occurs if X is a liquid (glides are absent from this position):

széppróza	[se:ppro:zɔ]	'fiction'
legkritikusabb	[lɛkkritikušɔbb]	'most critical'
két tragédia	[ke:ttrɔge:diɔ]	'two tragedies'

Summary

When the geminate is underlying or is produced by a morpho-phonological, ie. early lexical rule, Deg is obligatory. Types (1), (2) and (3a) belong here. In these cases it is immaterial what the flanking consonant is, the rule is triggered by any

consonant. When,however, the geminate is produced by later merger ("gemination") of two adjacent consonants, Deg is obligatory next to an obstruent (Types (3bi) and (4i)) but inapplicable next to a liquid or glide (Types (3bii) and (4ii)), and optional next to a nasal. Since affricates do not undergo such "gemination", they are not subject to such Deg either (kulcscsináló).

All in all: geminates are found adjacent to another consonant on the surface, but only if that consonant is a liquid or glide (or optionally nasal) and if the geminate is derived at some not-too-early stratum - typically, postlexically. Note that the geminates spared by these rules may fall victim to the fast-speech degemination that affects all geminate (or long) consonants of Hungarian.

References

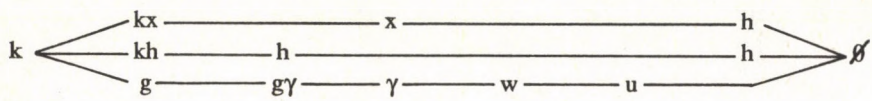
1. HALL, R.A.Jr.: Hungarian Grammar. Language Monograph No. 21, LSA, Baltimore, 1944.
2. MOHANAN, K.P.: The Theory of Lexical Phonology. Dordrecht, 1986.
3. PAPP, I.: Leíró magyar hangtan. Budapest, 1966.
4. VAGO, R.M.: The Sound Pattern of Hungarian. Washington, 1980.

SONORIZATION AND SPIRANTIZATION: A SINGLE PHONETIC PROCESS?

Ailbhe Ní Chasaide
Centre for Language and Communication Studies, Trinity College, Dublin 2, Éire

1. INTRODUCTION

Sonorization and spirantization are frequently treated as manifestations of a single phonological process, termed lenition. Sonorization (henceforth Vo) involves the change of consonants in the direction voiceless → voiced, which tends to occur in VCV. Desonorization, i.e. change in the opposite direction, voiced → voiceless, tends to occur in #CV and VC# environments. It is regarded as a fortition or strengthening, and these last environments characterized as 'strong'. Spirantization (henceforth W) involves the gradual weakening (or opening) and eventual loss of the supraglottal articulation of consonants e.g., k → x → h → ∅. It is most likely to occur in VCV, but can also be found in VC#; it is least likely to occur in #CV. Any change in the opposite direction is also regarded as a strengthening. Foley [1] further points out that voiced stops tend to undergo W before voiceless, velars before bilabials or alveolars, and that geminates show most resistance to W. The following lenition schema involving both processes is suggested by Lass and Anderson, [3]:



Given the quasi-universal status of lenition processes, it would be reasonable to expect an ultimately phonetic causation, i.e. that they are triggered by some constraints on human production or perception mechanisms. But in what sense are W and Vo related? Are they two distinctly different processes which just happen to occur in (note, not fully) overlapping environments, or do they share a common phonetic content? On the basis of many known and some new experimental findings, this paper suggests that these processes are indeed highly related and attempts to sketch what their phonetic bases might be.

2. LENITION AS TARGET UNDERSHOOT

The proposed account is in terms of target undershoot. This is in many ways in line with suggestions in earlier work (e.g., Lindblom, [4]). It differs however in that 'target' is not simply deemed to be an articulatory goal, but may involve the attainment of finely balanced articulatory and aerodynamic conditions, both of which are subject to temporal constraints. Undershoot implies failure to fully meet either of these prerequisites of a target. It results from the interaction of (a) the inherent instability of particular targets in specific positions, and (b) a triggering environment involving rapidly uttered, relatively unstressed syllables.

(a) All targets are not equal in the demands they make of the speech production mechanism. Those targets which demand more are deemed inherently unstable relative to those whose execution would seem 'easier' in production terms. Positional variation, i.e. the position in which a segment occurs within an utterance may affect articulatory, aerodynamic or temporal aspects of a target's execution. It seems best therefore to define a target's instability in terms of positional environment, and #CV, VCV and VC# will be considered here.

(b) The tempo and the amount of stress with which individual syllables are uttered are constantly being modulated in running speech. I would suggest that rapidly uttered unstressed syllables provide the further triggering environment for lenition of targets, affecting to the greatest degree those targets which have the greatest inherent instability. The two charac-

teristics of the triggering environment crucial for lenition are temporal compression and lack of stress. The consequences of temporal compression are obvious; if the time window of the segment is reduced, there may simply not be enough time to meet the target's articulatory/aerodynamic requirements. The presence or absence of stress is likely to be manifest ar all levels of speech production, and all may be relevant to lenition. Unstressed tokens appear to involve lower respiratory effort, so that subglottal pressure (P_S) and the potential rate of airflow through the vocal tract is lower (Ladefoged, [2]). The supraglottal articulation of vowels in unstressed syllables also seems to involve less muscular effort, and there are indications (though less conclusive) that this may also be true of consonantal articulation (see Tuller et al, [7]). And crucial to the account of Vo lenition below, at the laryngeal level it seems that a lesser degree of vocal fold abduction characterizes voiceless segments in relatively unstressed syllables (Ní Chasaide, [5]).

3. W: WEAKENING OF SUPRAGLOTTAL ARTICULATION

(a) *Basic factors determining target instability*

Articulatory: The articulatory target involves a closing/opening cycle involving antagonistic muscular gestures, which leaves it vulnerable to temporal compression; if these gestures overlap excessively they will partially cancel each other out and the articulatory target will not be attained. Least affected should be those targets for which the articulators are more 'mobile' or 'independent'. As the lips and tongue tip are more mobile (i.e. can move more rapidly) than tongue front or back, the closing/opening gestures can be executed in less time. Furthermore, as tongue tip and labial consonant articulations involve relatively independent articulators from those of vowels, they are likely to allow more temporal overlap with adjacent vowel articulation. It follows that velars and palatals are inherently more unstable than labial and tongue tip articulations, being more vulnerable to temporal compression. Geminates and stops adjacent to homorganic nasals are least vulnerable, having a longer interval in which to execute the supraglottal articulation. Positional variation, being a major determinant of the time allocation of the segment must be a major factor predisposing to W lenition: stops in VC# are considerable longer than those in #CV, which in turn are longer than those in VCV. Finally, the tendency for voiced segments to be shorter than their voiceless counterparts leaves them comparatively more susceptible to W through temporal compression.

Aerodynamic: This is likely to be important at the fricative stage of W lenition, particularly of voiced fricatives. A successful fricative depends on sufficient constriction and a sufficient airflow through that constriction to produce turbulence. Glottal adduction for the voiced fricative greatly reduces the volume velocity of flow through the supraglottal constriction, and this affects the intensity (and hence perceptibility) of oral frication. Failure to satisfy the aerodynamic requirement of fricative production is most likely to present a problem in any environment where P_S drops, e.g., the unstressed triggering environment. Westbury and Keating [8] further suggest that P_S tends to fall in utterance final position.

(b) *Triggering Environment*

Temporal compression: To the extent that this is a causative factor of W, one can predict on the basis of the above that: stops in VCV should be most prone to W lenition, those in VC# least. Voiced segments should lenite before the longer voiceless ones, palatals and velars should lenite before bilabial and tongue tip segments. Geminates should shorten rather than spirantize when subjected to tempo increases.

Destressing: This could further affect the stop fricative/targets in two ways. Firstly, any reduction in muscular effort at the supraglottal articulatory level constitutes in itself a form a W lenition (note that experimental results have not conclusively demonstrated such a reduction). Reduction in respiratory effort and P_S levels would affect the intensity and perceptibility of frication. Voiced fricatives are likely to be particularly vulnerable in this respect, as would utterance final positional variants, for the reasons outlined above.

4. VO: THE VOICING OF VOICELESS SEGMENTS

Phonatory targets may be the initiation, maintenance or cessation of vocal fold vibration. In each case, appropriate articulatory and aerodynamic conditions must be met, both of which are affected by the temporal constraints pertaining to a segment.

(a) *Basic factors determining target instability*

Articulatory: The articulatory gesture is context dependent. For the voiced stop, the vocal folds must attain (#CV) or maintain (VCV and VC#) an adducted state. The voiceless stop in VCV presents the most complex articulatory (laryngeal) gesture, requiring an abduction/adduction cycle. In VC#, appropriately timed abduction alone, or in #CV, adduction alone is all that is required, even though these stops do tend to retain traces of the opening/closing cycle (Ní Chasaide, [5]). (Due to space limitations, discussion here will not cover glottalized stops.) Muscularly controlled vocal fold movement is rather slow when compared to that of supraglottal articulators (Roach, [6]). For the –voice target in VCV, a number of factors conspire to render it most vulnerable to temporal compression: the relative complexity of the laryngeal gesture, the comparative slugishness of laryngeal muscles and the intrinsic brevity of this positional variant.

Aerodynamic: To attain a +voice or a –voice target, it is not enough to bring the vocal folds to an appropriate configuration. To sustain +voice, appropriate aerodynamic conditions must be met, i.e. a transglottal pressure drop (ΔP_{tg}) of about 2 cm H_20. The greater the degree of oral occlusion of a segment, the more a buildup in oral pressure will neutralize the ΔP_{tg} and lead to devoicing. Thus fricatives devoice less readily than stops, and [h] less readily than the other fricatives. The duration of a segment is also crucial: Westbury and Keating, [8], estimate that 'passive' (aerodynamically occasioned) devoicing will occur after about 60 ms closure in the voiced stop, barring additional physiological adjustments to facilitate voice maintenance. They also estimate that voice initiation in #CV demands greater respiratory effort and a ΔP_{tg} of 4 cm H_20. Furthermore, if one allows for a falling P_S in CV#, they suggest that 'passive' devoicing of the voiced stop will be accelerated, occurring at about 35 ms.

Effecting the cessation of vocal fold vibration is not purely a question of vocal fold adjustment either. As pointed out by Ní Chasaide, [5], the transition voice → voiceless may take up to 90 ms if the vocal tract is unoccluded. Even when devoicing is passively aided by vocal tract closure, some residual voicing is found in the voiceless stop. To sum up the aerodynamic aspects of +voice and –voice targets: it is easiest to voice (and most difficult to devoice) those segments with least supraglottal occlusion and shortest durations. Positional variation is relevant in that it affects segment durations, and in that the +voice target in #CV or VC#, would seem to require more respiratory effort than in VCV.

(b) *Triggering Environment*

Temporal compression: Shortening the time allocation of the segment is most likely to affect the articulatory (laryngeal) aspect of the voiceless target in VCV. If there is excessive overlap of the antagonistic abduction and adduction gestures, undershoot of the articulatory target will result. Furthermore, the reduced duration of stop closure disfavours a voiceless target in reducing the likelihood of aerodynamically induced devoicing.

Destressing: Destressing would appear to have consequences at the level of the laryngeal articulation and at the aerodynamic level. Photoelectric glottographic data reported in Ní Chasaide, [5], suggests that the vocal folds abduct to a lesser degree for voiceless consonants in relatively unstressed syllables. This reduction in the abduction target can not be simply explained in terms of reduced duration of unstressed segments (unstressed geminates are longer than stressed single stops but show considerably less glottal abduction). More recent (as yet unpublished) EMG data supports the contention that less muscular effort is involved in these unstressed tokens. The aerodynamic consequences of destressing should affect those targets which demand most in respiratory terms, i.e. the initiation of voice in #CV or its maintenance in CV#. To sum up, the failure to suppress fold vibration in VCV may result primarily from undershoot in an articulatory sense, partly because destressing leads to active reduction of the abduction target, partly because of excessive overlap of opposing gestures due

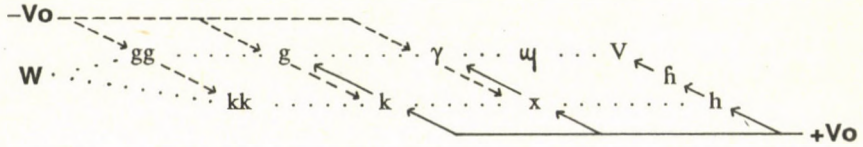
to to temporal compression. The failure to voice in #CV and VC# is more likely to represent aerodynamic undershoot, resulting from the inherent aerodynamic difficulty of these positional variants, and the reduced aerodynamic resources of the triggering environment.

5. CONCLUSION

This account argues in favour of treating W and Vo as a unitary process. Furthermore, it allows certain interactions between W and Vo to be predicted or explained. As we have seen, the greater the supraglottal occlusion, the more likely it is that voice will be suppressed, due to neutralization of the ΔP_{tr}. Thus, the further a segment has lenited along the W scale, $k \rightarrow x \rightarrow h$, the more susceptible it becomes to Vo. Given the inherent instability of voiced fricative targets (based on the fact that glottal adduction reduces the rate of flow through the oral constriction) further rapid W lenition would be predicted for these segments. Furthermore, any segment which has undergone Vo lenition is also likely to be more prone to W lenition. This is because voiced segments are shorter than their voiceless counterparts, and this should increase their susceptibility to W. Thus it can be said that W or Vo, once initiated, will tend to trigger each other.

The last stage of W lenition $h \rightarrow \emptyset$ is more likely to be $h \rightarrow \tilde{h} \rightarrow V$ and involve an instance of Vo rather than W. Lacking supraglottal occlusion, [h] is extremely susceptible to voicing, and disappears by being gradually absorbed into the adjacent vowel. The same is true of the last stage of W in voiced segments. Thus, W and Vo coalesce in their final stages.

The commonly occurring change voiced → voiceless in #CV and VC# have traditionally been treated as strengthenings, and the environments themselves regarded as 'strong'. This seems counterintuitive however; as it is precicely the +voice target which presents greatest difficulty to the production system, it would seem more reasonable to regard a change voiceless → voiced as constituting a strengthening in these environments. I would therefore suggest that Vo lenition should be treated as a bidirectional process, +Vo and −Vo. The directionality of change is determined by positional variation, as this dictates both the inherent instability of targets and the effects of the triggering environment. The interaction of W (dotted lines) with +Vo (solid lines) and with −Vo (dashed lines) is schematized as follows:



REFERENCES

[1] Foley, J. (1977): *Foundation of Theoretical Phonology*. Cambridge University Press.

[2] Ladefoged, P. (1967): *Three Areas of Experimental Phonetics*. Oxford University Press.

[3] Lass, R. and Anderson, J. (1975): *Old English Phonology*. Cambridge University Press.

[4] Lindblom, B. (1983): "Economy of speech gestures", in MacNeilage ed. *The Production of Speech*, New York: Springer Verlag.

[5] Ní Chasaide, A. (1985): "Preaspiration in Phonological Stop Contrasts", unpublished Ph.D. thesis, University College of North Wales, Bangor.

[6] Roach, P. (1978): "Reaction time measurements of laryngeal closure", Phon. Lab. Univ. Reading, Work in Progress 2, pp. 1-23.

[7] Tuller, B., Harris, K., and Scott Kelso, J. (1982): "Stress and rate: differential transformation of articulation", J.Acoust.Soc.Am., 71, pp. 1534-1543.

[8] Westbury, J.F. and Keating, P. (1980): "A model of stop consonant voicing and a theory of markedness", paper presented at Ann. Meet. Ling. Soc. Am., San Antonio, Texas.

STUDIES OF PHONOLOGICAL RULES IN THE SPEECH OF THE DEAF

Anne-Marie ÖSTER
Dept of Speech Communication and Music Acoustics,
Royal Institute of Technology, Box 70014, 100 44 Stockholm, Sweden

This paper presents a phonetic and phonological analysis of the speech of a 15 year old deaf boy who attends the 8th grade at the school for the deaf in Stockholm (the Manilla School). A speech material was recorded at intervals of three months to investigate to what extent deaf speech is rule-governed and stable over time. The result of the assessments shows the importance of a phonological analysis in addition to a traditional phonetic analysis for general descriptive purposes. A phonological analysis detects whether an error in the phonetic representation of the child's productions maintains a phonological contrast in his speech or not.

Introduction

During the last years we have been working on a project named "The speech in a bilingual deaf school". The aim of the project is to improve and develop efficient speech training methods for a new teaching situation in the schools of the deaf where sign-language is the first language and Swedish is the second language. This will mainly be learnt through written Swedish. A lot of work has been done on developing applicable methods for analyses and diagnoses, which should be the basis of an efficient and individualized speech training. The demand on such methods is to detect the possible existence of regular deviations, phonetic context of the deviations, and stability of the deviant phonological representation. Consequently, a traditional phonetic analysis that only describes distortions, substitutions, and omissions in the deaf child's speech is inadequate and should be supplemented with a phonological analysis that shows how the child uses his articulation in linguistic contexts.

The speech of congenitally deaf children is characterized by many errors which seem to be systematic. The assumption that deaf speakers have well-established speech habits, that is, a speech sound in the same phonetic context is always pronounced in the same way, seems to be true but has not yet been studied in detail. To study this assumption we recorded speech from a deaf adolescent boy reading the same material at an interval of three months. The speech samples from two recordings were analysed to investigate the stability between the two readings and his phonological representation of Swedish consonants.

Data collection

The child's speech was video recorded with an interval of three months. At both times the child read the same speech material that consisted of monosyllabic and disyllabic words as well as sentences which contained these words. The words were chosen to be common and familiar to the child. Some of the consonants did not occur in all possible Swedish positions, as is seen in Table 1.

Transcription

The video recordings were transcribed using the symbols of the International Phonetic Alphabet. Many peculiarities and fusions of errors occur in prelingually deaf children's speech, which make an expansion of diacritical marks necessary. We have used some of those which Bush & al. (1) and Roug & al. (2) have developed for the transcription of babbling and phonetic development in early infancy, see Fig. 1.

Results

Stability between two readings

The results of a narrow phonetic analysis of the two recordings are summarized in Table 1 according to the position in the word and to standard phonological representation.

The results presented in Table 1 show that there was a high stability in the child's consonant production and the phonological representation between the two readings. Each square is based on at least four readings and represents isolated words as well as words in sentences.

b̯	VERY SHORT	b̰	NASAL AIR EMISSION
bʰ	ASPIRATION	b̃	NASAL
b̥	DEVOICED	nj	COARTICULATION
b̪	LIP PROTRUSION	n̪	INTERDENTAL

Fig. 1. EXAMPLES OF DIACRITICAL MARKS USED IN THE ANALYSIS

IPA	INITIAL		MEDIAL		FINAL	
	I	II	I	II	I	II
/p/	b̯	b̯	bʰ	bʰ		
/t/	d	d	dʰ	dʰ	d̯•	d̯•
/k/	g•	g•	gʰ	gʰ	g̯•	g̯•
/b/	b̲	b̲				
/d/	d	d	dʰ	dʰ	d̯	d̯
/g/	g̰	g̰	gʰ	gʰ	g̯	g̯
/f/	f	f				
/v/	v̰	v̰, mv̰			v̰	v̰
/s/	- s ʂ	- s	s	-	s	s
/ʃ/	ʂ	s				
/j/	nj	j̰				
/h/	x	x				
/l/	l̰	l̰	d̯~	d̯~	l̰	l̰
/r/			r	r	r	r
/n/	n̪	n̪	n̪	n̪	n̪	n̪
/m/	m	m				

Table 1. Narrow phonetic analysis of the subject's two readings according to standard phonologic representation and to position in the word.

The only exception that showed a phonologic instability between the two readings was the use of the fricatives /s/ and /ʃ/ in initial and medial position. The child was very unsure which sound to pronounce in the first reading. Sometimes he used [s̪] , sometimes [s], and sometimes he simply omitted it. To him, there was no difference between the two phonemes. He used them as allophones of the same phoneme. In initial consonant combinations (e.g. *skr*att) he omitted /s/ in both readings. In the second reading, he reduced the system and produced [s] for both /s/ and /ʃ/ initially and omitted /s/ in medial position.

Phonological analysis

Table 2 shows the phonetic and the phonological system of the child's stops in the second reading.

The result gives some interesting information about the child's speech. The column to the left in the table is the result of a general phonetic analysis that lists the inventory of correctly produced Swedish consonants. According to this analysis, it seems obvious that /p/, /t/, and /k/ are missing in the child's speech and that he controls the production of the voiced stops.

COARSE PHONETIC LABEL	NARROW PHONETIC REPRESENTATION		PHONOLOGIC	
[b]	[b̪]	—	[p]	INITIAL POSITION
	[b̲]	—	[b]	INITIAL –"–
	[bʰ]	—	[p]	MEDIAL –"–
[d]	[d]	—	[d/t]	INITIAL –"–
	[dʰ]	—	[d/t]	MEDIAL –"–
	[d̥]	—	[t]	FINAL –"–
	[d̰]	—	[d]	FINAL –"–
	[d̃]	—	[l]	MEDIAL –"–
[g]	[gʰ]	—	[k]	INITIAL –"–
	[g̊]	—	[g]	INITIAL –"–
	[gʰ]	—	[g/k]	MEDIAL –"–
	[g̊]	—	[k]	FINAL –"–
	[g̰]	—	[g]	FINAL –"–

Table 2.
The phonetic and phonological representation of the child.

However, the columns to the right show the child's articulation errors for each sound individually and what the substitutes look like. The result from the phonological analysis shows that the child makes a distinction between /b/ and /p/ in initial position but not through a voice/voiceless contrast. He makes a distinction by lip-protusion.

Only in final word position does he make a contrast between dental stops by adding a neutral vowel after a very short, fully devoiced /d/. He does not control the voice/voiceless contrast between /d/ and /t/. In initial position, he produces a voiced dental stop for both /d/ and /t/, and in medial position, /d/ and /t/ are produced as an aspirated /d/.

The same strategy is used for the velar stops. Initially, the distinction is made by adding a neutral vowel sound and by nasal air emission. Medially, there exists no contrast between /g/ and /k/. Both are produced as an aspirated voiced velar stop. In final position of a word, the distinction between /g/ and /k/ is made by adding a neutral vowel to a fully devoiced, very short /g/ to represent /k/.

Principal phonological processes

The principal phonological processes that could be found in the child's stop production are consequently:

Voicing of voiceless stops in initial positions, aspiration and voicing of voiceless stops in medial positions, and devoicing of voiceless stops in final positions. In addition to that, regular error patterns of expressing speech sound contrasts are found in initial position between bilabial stops by lip-protrusion and in final positions between dental and velar stops by adding a neutral vowel.

Discussion

The analysis of our subject's speech supports the idea that the speech of the deaf is rule-governed. It is reasonable to assume that lip-reading, residual hearing, teaching methods, sign language, and writing, which are the effects of a bilingual teaching method, may impact on the phonological development of the deaf child's speech.

The analysis of his fricative and stop production in the two readings illustrates the importance of a phonological analysis, which shows how the child uses the speech sounds in a linguistic context. A traditional descriptive analysis had provided misleading information about the child's fricative production. The conclusion had been drawn that the child controlled the phonetic contrasts between /s/ and /ʃ/ because both sounds occurred in his sound system of the first reading. However, the fact that he did not control the phonological contrasts of the two sounds had not been detected. A phonological analysis also detects whether a deviant pronunciation may actually maintain a speech sound contrast as is seen in the child's stop production. At the same time, important information about regular error patterns may be derived from this analysis.

References

1. BUSH, C.N.--EDWARDS, M.L.--LUCKAU, J.M.--STOEL, C.M.--MACKEN, M.A.--PETERSEN, J.D.: On specifying a system for transcribing consonants in child language: A working paper with examples from American English and Mexican Spanish. Report, Stanford University, Dept. of Linguistics, 1973.
2. ROUG, L.--LANDBERG, I.--LUNDBERG, L-J.Phonetic development in early infancy. A study of four Swedish children during the first 18 months of life. Report, University of Stockholm, Dept. of Linguistics, 1987.

CONTRIBUTION TO UNIVERSAL CLASSIFICATION OF PHONEMIC CHANGES

Jacek PERLIN
Department of Neophilology
University of Warsaw, Warsaw, Poland

1. The basic principle of this investigation was the conviction that phonetic and phonological changes are classifiable and that this classification is useful. Obviously, it is possible to systematize all the phenomena, but their classification can have cognitive values only if they meet some requirements, first of all, if they are formally heterogenuous and do not present the same probability of occurrence. And so, if every sound could, with the same probability, turn, as a result of a phonetic change, into any other (even within a given class), then such a systematization would be deprived of greater value. However, observing phonetic changes in genetically and typologically different languages it is easy to note that some kinds of modification are very frequent, others less frequent and some probably do not occur at all. At the same time, many examples of bifurcation of languages that give first dialectal variants and then different languages prove that the evolution can make various ways, independent of any external influences and internal conditions. So, there is no doubt that there exist many possibilities of sound or word changes. It implies that a formulation of universalia of diachronic phonetics is, it seems, a priori impossible. Taking into consideration the impossibility of foreseeing change directions, one can only indicate theoretical variants of evolution, i.e. determine which changes are possible and which are not. In other words, eventual universalia could be only negative. Nevertheless, a classification of phonemic changes seems important not only as an objective for the sake of itself, but also as a contribution to further investigations with a view to settle relations between phonetical and phonological changes, i.e., the influence of sound changes on a system. To achieve this, a methodological and terminological apparatus is required, based on homogenuous, coherent criteria and subsequent assumptions.

2. Elaborating the classification criteria we have taken into consideration formal constitutional elements of words as quality and quantity of sounds; their relative length; quality of syllables and their boundaries; physical features and place of appearence of prosodical phenomena. With reference to the all above mentioned elements we have applied a trichotomeous opposition of quantity, quality and "positionality". As for functional changes, i.e. referred to a system functioning, we have assumed that the phonemic system is composed of three levels: phonetic, phonological and prosodic. A syllable was treated separately as a specific unit. Functional changes have been divided, as formal ones, into quantitative, qualitative and positional.

3. The elaborated classification has the following structure:

I. Quantitative changes

A. Segmental quantitative-formal changes
- phonic reduction (reduction of word constitutional elements)
- phonic increase (increase of word constitutional elements)
- articulatory reduction (reduction of the average time of word articulation)

- articulatory increase (increase of the average time of word articulation)
- syllabic reduction (reduction of the number of syllables in a word)
- syllabic increase (increase of the number of syllables in a word)

B. Quantitative-functional changes
- disphonologization (reduction of the number of phonemes of a system)
- phonologization (increase of the number of phonemes of a system)
- disphonetization (reduction of the number of allophones of a system)
- phonetization (increase of the number of allophones of a system)
- disdiacritization (reduction of the quantity of diacritic features of a system)
- diacritization (increase of the quantity of diacritic features of a system)

C. Distribution changes
- phonological exclusion (phoneme becomes impossible in a given position/environment)
- phonological access (phoneme becomes possible in a given position/environment)
- phonetic exclusion (allophone becomes impossible in a given position/environment)
- phonetic access (the opposite)
- decrease of phoneme productivity/frequency
- increase of phoneme productivity/frequency
- decrease of allophone productivity/frequency
- increase of allophone productivity/frequency

D. Quantitative-formal prosodic changes
- prosodic reduction (reduction of the number of prosodemes of a system)
- prosodic increase (the opposite)

E. Quantitative-functional prosodic changes
- disaccentization (disappearence of the accent/stress in a system)
- accentization (appearence of the/a tone in a system)
- tonalization (appearence of the/a tone in a system)

F. Quantitative-functional prosodic changes based on distribution
- accent exclusion (accent/stress becomes impossible in a given syllable)
- accent access (the opposite)
- tone exclusion (tone becomes impossible in a given syllable)
- tone access (the opposite)

G. Formal syllabic changes
- shift of syllable limit

H. Syllabic structure changes
- differentiation of centres (syllables in different positions begin to differ in the inventory of sounds that can form their centres)
- unification of centres (the opposite)
- paradigmatic centre change in consequence of exclusion
- paradigmatic centre change in consequence of access
- reduction of maximum syllabic slope (in a system)
- increase of maximum syllabic slope
- paradigmatic slope change in consequence of exclusion (given element becomes impossible in the slope of a syllable)

- paradigmatic slope change in consequence of access (the opposite)

J. Changes of syllable functions
- prosodization (syllable becomes carrier of prosodic features)
- disprosodization (the opposite)

II. Qualitative changes

K. Articulatory quantitative changes
a) general terms: opening, centralization, diphthongization..etc.
b) minute terms: dorsalization, dentalization, nasalization, disnasalization, sonorization...etc.

L. Quantitative-formal prosodic changes
- lenghthening,fortition, melodizing, discomposition..., etc.

M. Qualitative-functional prosodic changes
- loss of culminative function
- acquirement of culminative function
- loss of delimitative function
- acquirement of delimitative function
- disdistinctivization of accent place
- distinctivization of accent place

III. Positional changes

N. Segmental formal positional changes
- interversion (change of place of contiguous sounds)
- metathesis (change of place of discontiguous sounds)

P. Positional prosodic changes
- change of accent place in a word

Q. Functional-positional changes
- transphonologization (change of relations within the inventory of phonemes)

4. The above-mentioned classification shuld enable the description of all sorts of phonemic changes, not only testified but also possible to conceive. It does not seem, however, the classification contains superfluous elements that could not serve to describe real linguistic phenomena; anyway, for a great majority of the change categories it was not too difficult to find convincing examples from the history of any well described language.

5. The principal employment of the above classification could be its application in the analysis of relations between sound and system changes. The most interesting observation which results from the examination of rapports between formal and functional changes are the following: phonologization, i.e. an increase in a phoneme inventory is practically always a consequence of a reductive change on the formal level; prosodic changes do not influence the phonological and distributional systems; transphonologization can be only an effect of qualitative changes; reductive changes influence much more the phonological and distributional systems than epenthetic ones.

UNDERSPECIFICATION THEORY AND VOWEL HARMONY

Catherine O. RINGEN
Department of Linguistics
University of Iowa, Iowa City, IA, 52242, USA

This paper considers vowel harmony (vh) in Finnish, Hungarian, Igbo, and Kalenjin within the general framework of Underspecification Theory (UT). It is suggested that certain exceptions to vh in these languages are not easily accounted for in UT unless [+F], [-F], and [F] (unspecified for F) are permitted in the same context in underlying representations and throughout derivations.

In native Finnish non-compound words, front harmonic vowels (ü=[y], ö=[ø], ä=[æ]) do not occur in words with back harmonic vowels (u, o, a); the neutral vowels (i, e) occur in words with either front or back harmonic vowels. Only in loans, compounds, and slang words, do front and back harmonic vowels co-occur. Harmonic suffix vowels alternate depending on the harmonic quality of root vowels (e.g. pää-llä 'head' adess., maa-lla 'land' adess., lapse-lla 'child' adess., järve-llä 'lake' adess.). If all the vowels in a root are neutral, harmonic suffix vowels are usually front (e.g. tie-llä 'road' adess.).

If it is assumed, following Goldsmith (1985), that Finnish VH is a rule that spreads [-back] rightward, then these data are easily described in UT. Vowels can be assumed to be specified as in (1):

(1)

	i	e	u	o	a	y	ö	ä
high		-		-			-	
low					+			+
back						-	-	-
round	-	-						

The Redundancy Rules (RRs) in (2) fill in the unspecified values, but do not change features.

(2) a. $V_{[+low]} \rightarrow \begin{bmatrix}-high\\-round\end{bmatrix}$ d. $V_{\begin{bmatrix}-round\\-low\end{bmatrix}} \rightarrow [-back]$

b. $V \rightarrow [+round]$ e. $V \rightarrow [+back]$

c. $V \rightarrow [-low]$ f. $V \rightarrow [+high]$

Following Pulleyblank (1986), I assume that unassociated autosegments are linked by the Universal Association Convention (UAC) to unassociated vowels from left to right, one to one.

Forms such as pää and tie are assumed to have unassociated [-back] autosegments which the UAC associates with the leftmost vowel. VH then applies to associate this [-back] segment with the remaining root vowels and suffix vowels. In the case of back vowel roots such as maa, there is no underlying specification for backness; both root and (harmonic) suffix vowels receive [+back] specification by the RRs which apply as late as possible. The underlying form of lapse-lla would also be unspecified for backness; the first and last vowels, originally specified only as [+low], would undergo the RRs in (2) and surface as a. The second vowel, specified only as [-round, -high], would, by application of the RRs, surface as e. There is one complication which should be noted here. According to the Redundancy Rule Ordering Constraint (RROC) (Archangeli, 1984), the RR (2d), which fills in [-back], will be assigned to the same component as VH. The RROC states: A RR assigning [αF], where "α" is "+" or "-", is

automatically assigned to the first component in which there is a rule which refers to [αF]. The rule (2d) does not actually apply, however, because the vowel e is not yet specified as [-low] and hence the structural description of the rule is not met. Were (2d) to apply to the vowel e in lapsella, the analysis sketched here would not work because subsequent application of VH would incorrectly spread the [-back] to the suffix vowel. Thus, (2d) will not apply until after (2c) has applied, which will be after VH has ceased to be applicable. (It is assumed in UT that once a RR has been activated, it applies whenever its structural description is met.)

In disharmonic loans, suffix vowels generally agree with the last harmonic root vowel, e.g. afääri-llä 'affair' adess., syntaksi-lla 'syntax' adess. The underlying form of syntaksi-lla can be assumed to have a [-back] autosegment which is lexically bound to the first vowel. It is usually assumed that the Strict Cycle Condition (SCC) blocks the root internal spreading of such a lexically bound autosegment. The idea is that cyclic rules apply on a given cycle only to derived representations. On the first cycle, [-back] may not spread to the other root vowels because the representation is not derived (the structural description is met by the lexical representation) and, hence, root internal spreading is blocked. In general, however, root internal spreading is not blocked because the structural description of VH is not met by the lexical representation; application of the UAC creates a derived input to which VH applies. The SCC will not, however, block the spreading of the [-back] to the suffix vowel of syntaksi-lla. Recent discussions of autosegmental phonology have assumed that all spreading is strictly local. Archangeli and Pulleyblank (1987), adopt the Locality Condition (LC), which states that a rule can apply only if a specified target is adjacent to a specified trigger. They note that if features are hierarchically organized, as suggested by Clements (1985), and if a rule whose target is node or feature α scans the highest level of syllabic structure providing access to α, (maximal scansion) then, in general, consonants will be transparent to rules affecting vowels, but vowels will block rules applying to consonants. This is because rules whose targets and triggers are vowels will scan at the level of syllable heads, a level which provides access to vowels, but not consonants, whereas rules affecting consonants will scan the skeletal tier, a level which provides access to both consonants and vowels. The LC will block the spreading of [-back] of syntaksi to the suffix vowel because the target and trigger are not adjacent vowels. The treatment of afääri-llä is more problematic. If we assume that like syntaksi, afääri has a [-back] autosegment bound to ä, then for the same reasons that the [-back] autosegment of syntaksi cannot spread, the [-back] autosegment of afääri cannot spread and the suffix vowel will incorrectly become back. The most straight forward solution is to assume that the underlying representation of afääri has a [+back] and a [-back] autosegment. By the UAC these autosegments are associated with unassociated vowels, one to one, left to right. The [-back] would not be blocked by the SCC or by the LC from spreading to the remaining root vowels or to suffix vowels, and the correct form would be derived. Notice, however, that this analysis requires that [+back], [-back], and [back] occur in underlying representations, a possibility that has been explicitly rejected by proponents of UT.

A description of vh in Hungarian, almost identical to the one just outlined for Finnish, can be given in UT (see Ringen 1988). Hungarian disharmonic loans such as manöver 'maneuver' seem to show that [+back], as well as [-back] and [back] are needed in an account of Hungarian for the same reasons given above for the parallel example of afääri in Finnish.

Consider next the case of Kalenjin which has asymmetric or dominant-recessive vh. The vowels of Kalenjin can be divided into two sets, those that are [+ATR] ([+Advanced Tongue Roots]) i, u, e, o, ɒ and those that are [-ATR], ɪ, ʊ, ɛ, ɔ, ɑ. Morphemes in Kalenjin fall into two classes, those that alternate and those that do not. For the most part, morphemes that do not alternate have [+ATR] vowels, whereas alternating morphemes have either [+ATR] or [-ATR] vowels. When a word is made up entirely of alternating morphemes, all the vowels of the word are [-ATR]. When a non-alternating morpheme with [+ATR] vowels occurs anywhere in a word, whether it is prefix, suffix, or root, all vowels in the word are [+ATR]. kɪ-ɑ-gɛr-∅ (distant past-I-shut-it) 'I shut it', is made up entirely of alternating morphemes and all vowels are [-ATR]. ki-ɒ-ge:r-in (distant past-I-see-you sg. obj.) 'I saw you', in contrast, contains the non-alternating root 'see' and all vowels are [+ATR]. ki-ɒ-ger-e 'I was shutting it' contains the non-alternating noncompletive suffix -e and all vowels in the word are [+ATR].

The Kalenjin data can be analyzed within the framework of UT by assuming that non-alternating morphemes have [+ATR] autosegments while alternating morphemes are unspecified for ATR. VH a rule that spreads [+ATR] bidirectionally. Vowels which remain unspecified for ATR are specified as [-ATR] by RR. Thus, the underlying representations for ge:r 'see' and -e, the noncompletive marker, would have [+ATR] autosegments, the forms gɛr/ger 'shut', kɪ/ki 'dist. past., and ɑ/ɒ 'I' would be unspecified for ATR. In this analysis, the [-ATR] vowels in kɪ-ɑ-gɛr result from a RR. The [+ATR] vowels in ki-ɒ-ger-e result from association of the [+ATR] autosegment of the suffix -e by the UAC and subsequent application of VH.

There are three non-alternating morphemes with [-ATR] vowels that never have [+ATR] vowels, even in a word with [+ATR] vowels. These are the negative prefix mɑ -, the perfectivizer kɑ ~ gɑ, and the reflexive suffix kɛ: ~ -gɛ: which occur in the following examples: ki-ɒ-un-gɛ: 'I washed myself', ma-ti-un-gɛ: 'don't wash yourself', kɑ-mɑ-ɒ-ge:r-ɒk 'I didn't see you (pl.)' and kɑ-mɑ-gɑ-go-ge:r-ɒ 'and he hadn't seen me'. The last two forms show that not only do the non-alternating morphemes fail to become [+ATR], they also block the spread of [+ATR] to the regularly alternating suffix vowels to their left. These exceptional affixes, like exceptional Finnish and Hungarian roots, seem to require that [-ATR], as well as [+ATR] and [ATR], occur in lexical representations.

Igbo has symmetric vh that is superficially different from Kalenjin, but which has a VH rule that also spreads [+ATR] (see Ringen, 1979 and to appear). Unlike Kalenjin, however, Igbo has no affixes with underlying [+ATR]. Igbo also has non-alternating affixes which are [-ATR] and which block the spread of [+ATR] to regularly alternating affixes. These forms also seem to require that [-ATR], like [+ATR] and [ATR], occur in lexical representations (see Ringen, to appear).

The suggestion that [+ATR], [-ATR], and [ATR] occur in phonological representations and throughout derivations does not necessarily entail that features are ternary and not binary. Obviously, if representations contain [+F], [-F] and [F] (unspecified for F), and if it is possible to refer in rules to [F], as well as to [+F] and [-F], the system is ternary, not binary. The well-known Lightner-Stanley argument shows that an unspecified value can function as a third value distinct from '+' and '-', even if reference to the unspecified value is not permitted. If, however, reference to unspecified values is prohibited and if the RROC is adopted, then unspecified values do not function as third values and features are binary not ternary.

References

1. ARCHANGELI, D.: Underspecification in Yawelmani Phonology and Morphology. MIT dissertation, 1984. Published, New York, N.Y., 1988.
2. ARCHANGELI, D.--PULLEYBLANK, D.: Maximal and minimal rules: effects of tier scansion. Proceedings of NELS 17. GLSA, University of Massachusetts, Amherst. 1987, 16-35.
3. CLEMENTS, G.: The geometry of phonological features. Phonology Yearbook 2. 1985, 225-252.
4. GOLDSMITH, J.: Vowel harmony in Khalkha Mongolian, Yaka, Finnish, and Hungarian. Phonology Yearbook 2. 1985, 251-275.
5. RINGEN, C.: Vowel harmony in Igbo and Diola Fogny. SAL 10. 1979, 247-259.
6. RINGEN, C.: Transparency in Hungarian vowel harmony. Phonology 5.2. 1988, 327-342.
7. RINGEN, C.: Underspecification theory and binary features. To appear in H. van der Hulst and N. Smith (eds.). Features, Segmental Structure and Harmony Processes. Dordrecht: Foris, to appear.

HOW MANY AFFRICATES ARE THERE IN HUNGARIAN?

Péter SIPTÁR
Institute of Linguistics,
Hungarian Academy of Sciences, Budapest, Hungary

The question raised in the title appears to be rather elementary, yet the correct answer is far from being obvious, let alone generally accepted. Possible answers range between two (/t͡s/, /t͡ʃ/) and as many as twelve (/t͡s/, /t͡s:/, /d͡z/, /d͡z:/, /t͡ʃ/, /t͡ʃ:/, /d͡ʒ/, /d͡ʒ:/, /c/, /c:/, /ɟ/, /ɟ:/); what is more surprising, both extremes, as well as most intermediate numbers, have actually been suggested as the correct answer in the literature. To be sure, all twelve items do appear as surface phonetic segments in Hungarian speech; which ones are to be granted phonemic status is the issue considered in this paper.

It is easy to see that the six long candidates can be explained away as geminates and/or (fused) clusters. The remaining six items fall into three classes in terms of whether their interpretation raises problems and, if it does, what types of problems are involved. In particular, /t͡s/ and /t͡ʃ/ are uncontroversial: they are definitely affricates in terms of their phonetic makeup, and phonologically they are obviously independent (monophonematic) members of the inventory of phonemes. Their voiced counterparts, [d͡z] and [d͡ʒ] are also undoubtedly affricates but their monophonematicity is less obvious. Finally, /c/ and /ɟ/ represent the opposite case: there is no doubt as to their phonemic status, but they may be interpreted either as palatal affricates or as palatal stops, depending on which of their surface realizations — both types being attested — are taken to be basic. Let us start with the latter issue.

The surface realization of the two palatal obstruents may be affricate-like to a variable extent, depending on phonetic context. Before stressed vowels (*tyúk* 'hen', *gyár* 'factory') and word finally (*fütty* 'whistle', *vágy* 'desire'), they are quite strongly affricated; before an unstressed vowel — especially for /ɟ/ as in *magyar* 'Hungarian' — much less, and before an oral stop (*ágyba* 'to bed') not at all. The fricative component is usually absent before /r/ (*bugyrok* 'bundles'); before /l/ lateral release can be observed as with stops (compare *fátylak* 'veils' with *hátlap* 'reverse side'), and only under strong emphasis do we find a fricative component as with true affricates (cf. *vicclap* 'comic journal'). Of the nasals, /m/ may be preceded by slight affrication (*hagyma* 'onion'), but /n/ and /ɲ/ may not (*hagyna* 'he would leave some', *hegynyi* 'as large as a hill'). The degree of affricatedness depends further on style and rate of speech: in slow, deliberate speech it is much stronger than in fast or casual styles. This wide range of variables and varieties should raise our suspicion that we have basically stops here which, under the appropriate circumstances, get more or less affricated due to well-known physiological factors; notice that true affricates do not exhibit such extensive variability. Consider English /t/ as an analogous case: in some dialects and in some environments it is affricated into [tˢ] — but this obviously does not affect its place in the consonant system of English.

Now to a more specific type of argument. Stops can be realized by their non-released allophones before another stop, e.g. *kapta* [kɔp°tɔ] 'he got it', *rakta* [rɔk°tɔ]

'he put it', whereas affricates obviously cannot, since they do not have such allophones: *bocskor* [botʃkor] (*[bot̚°kor]) 'moccasin' *barack* [bɔrɔt͡sk] (*[bɔrɔt°k]) 'peach'. Now /c ɟ/ are usually unreleased in this position: *hegytől* [hɛc°tøl] (*[hɛc̚ɕtøl]) 'from the hill', *hagyd* [hɔɟ°d] (*[hɔɟ̚ɟd] 'leave it'; in some cases (before velars?) there is vacillation: *hetyke* [hɛc°kɛ] (~[hɛc̚ɕkɛ]) 'pert'. This property clearly shows that they pattern with stops. As a corroboration, consider the related fact that affricates are less prone to LCF (long consonant formation) across word boundary than stops are, cf. *Gács Csaba* vs. *Tóth Tamás*. Now if we look at phrases like *ramaty tyúk* 'decrepit wench', *nagy gyár* 'big factory', we find that LCF applies automatically and obligatorily — as it is expected for stops, as opposed to true affricates. This should not come as a surprise, given that a geminate stop is nothing else but a sequence of an unreleased and a normal allophone of the same stop consonant.

In sum: /c ɟ/ are palatal stops in Hungarian; in the appropriate phonetic context, under appropriate conditions in terms of stress, speech rate, and speech style, they get affricated, as is to be expected for physiological reasons and can be observed in other languages that have palatal stops.

Turning now to [d͡z], [d͡ʒ]: here we have to consider if these are monophonematic affricates like [t͡s], [t͡ʃ], or stop + fricative clusters.

The speech sound [d͡z] can come from three sources in Hungarian. It can be a voiced allophone of the phoneme /t͡s/ (*lécből* [le:d͡zbøl] 'out of lath', *táncba* [ta:nd͡zbɔ] 'into the dance'), where obviously no underlying /d͡z/ is involved. It can occur in words like *pénz* [pe:nd͡z] 'money', *benzin* [bɛnd͡zin] 'petrol'; here, however, we have /nz/ clusters where [d] is an inorganic, epenthetic segment like [p] in *szomszéd* 'neighbour', [b] in *oromzat* 'gable', [c] in *München* 'Munich', etc. Finally, in words like *madzag* 'string', *bodza* 'elder', *pedz* 'nibble', [d͡z:] can be analysed in one of two ways (accepting the geminate analysis of long consonants): either as geminate /d͡zd͡z/→[d͡z:], cf. *vicces* 'funny' /t͡st͡s/→[t͡s:], or as /d-z/→[d͡z:], cf. *játszik* 'he plays' /t-s/→[t͡s:]. The first solution would involve positing a phoneme /d͡z/.

But this phoneme would have a rather skewed distribution: it would not occur word initially or postconsonantally at all; preconsonantally it would occur in a handful of suffixed forms; whereas intervocalically and finally (between vowel and word boundary) it would only occur doubled (long). This peculiar distribution, not found for any other member of the Hungarian consonant inventory, would be automatically explained by the cluster analysis (assuming an independently motivated realization rule converting a cluster of stop + sibilant into a long affricate). Let us consider what can be brought up against such an analysis.

Three types of possible counter-arguments come to mind. (a) The surface contrast between long affricates as in *madzag* 'string' and [d] + [z] clusters as in *vadzab* 'wild oats' shows that the former cannot be derived from an underlying cluster. (b) $C_iC_jC_k$ clusters (e.g. *kardvirág* 'cornflag') do not generally get simplified (fast-speech deletions aside), whereas $C_iC_iC_j$ clusters (e.g. *keddre* 'by Tuesday') do. Given that a stem-final (long) *dz* is shortened before a consonant-initial suffix, it follows that it cannot be a cluster. (c) Words like *vakaródzik* 'scratch oneself' can have short intervocalic [d͡z]; this makes the distribution less skewed and the independent phoneme analysis more plausible. — Are these three counter-arguments valid?

(a) The phonetic difference between *madzag* 'string' ([d͡z:]) and *vadzab* 'wild oats'

([d-z]) is totally parallel to that between *metszi* 'he cuts it' ([t͡s:]) and *hátszél* 'tail-wind' ([t-s]): in *vadzab/hátszél* internal boundary (compound boundary) occurs between stop and fricative, and it is that boundary that blocks their coalescence into a single long affricate. Hence, any counter-argument based on surface contrast of the *madzag/vadzab* type is unfounded.

(b) Next to another consonant in the word domain, all Hungarian long consonants get shortened (*sakktól* [ʃɔktol] 'from chess', *érvvel* [e:rvɛl] 'with argument'): this applies to [d͡z:] as well (*edzve* [ɛd͡zvɛ] 'being trained'). This, however, only proves that the immediate input to degemination is [d͡z:] (rather than a cluster); what it does not prove is that that [d͡z:] should go back to /d͡zd͡z/ and not /d-z/. Hence, this counter-argument fails, too.

(c) In words like *vakaródzik* 'scratch oneself', there is free variation (for some speakers) between short [d͡z] and long [d͡z:] (as well as simple [z]). This seems to refute our claim above, i.e. that there are no intervocalic short [d͡z]'s. But free variation proves exactly that length is irrelevant in this position: in other words, no short:long opposition is possible here. Since in non-vacillating cases (*madzag*) it is always long [d͡z:] that occurs, it is quite easy to see that in words like *vakaródzik* the segment in question is not short /d͡z/ but a long [d͡z:] whose actual length varies (tends to get reduced in long words like this); this [d͡z:], in turn, may just as well go back to a /d-z/ cluster. Hence, all three potential counter-arguments have turned out to be cases that can be easily accounted for in terms of the cluster analysis, too.

The existence of /d͡z/ as a phoneme, therefore, is not supported by any valid argument at all. The case of [d͡ʒ], however, is different in that arguments for /d͡ʒ/ are more or less balanced by arguments for /d-ʒ/. Word initial occurrence (as in *dzsámi* 'a type of mosque', *dzsóker* 'Jolly Joker') points toward /d͡ʒ/), whereas the behaviour of word internal [d͡ʒ]'s is practically identical with that of [d͡z], thus supporting a /d-ʒ/ analysis. This ambiguity could be resolved, in principle, in three different ways.

1. We could assume that — obviously with the exception of assimilation cases like *rácsban* [ra:d͡ʒbɔn] 'in grating' — [d͡ʒ] always goes back to a /d-ʒ/ cluster. In this case, the scope of degemination should be extended to include word initial position. Since word initial geminates are impossible anyway, such a redundancy rule (morpheme structure condition or surface phonetic constraint) is needed in any case — it should simply be allowed to operate during a derivation in which an offending representation is created by the coalescence of /d-ʒ/ into [d͡ʒ:].

2. Another possibility would be to claim that *dzsámi* 'jami' is /d͡ʒa:mi/ with an underlying affricate but *hodzsa* 'hodja' is /hod-ʒɔ/ with a cluster; this would explain the ambiguity referred to above but would give /d͡ʒ/ a rather skewed distribution (and it would be impossible to tell whether words like *lemberdzsek* 'anorak' should be taken to contain an underlying affricate or a /d-ʒ/ cluster).

3. Finally, we could accept the view that [d͡ʒ] is monophonematic everywhere; but then it is to be explained why its intervocalic (*menedzser* 'manager') and final (*bridzs* 'bridge (card game)' occurrences are invariably long (with a few exceptions like *fridzsider* [-id͡ʒi-] 'refrigerator' or *Roger Moore* [-od͡ʒɛ-]). It might be suggested that a kind of loanword gemination is at work here (cf. *dopping* /-pp-/ 'doping', *szvetter* /-tt-/ 'sweater', *sakk* /-kk/ 'chess', *meccs* /-t͡ʃt͡ʃ/ 'football match'). This looks quite feasible for items like *menedzser* and *bridzs*; the trouble is that the layer of vocabulary includ-

ing e.g. *hodzsa* 'hodja' does not exhibit this process, cf. *mecset* (**meccset*, **mecsett*) 'mosque', etc.

The first solution is technically neat and logically coherent; unfortunately, it does not conform to speakers' intuition and is rather abstract. What is more serious, /d- / as an initial cluster does not fit the overall pattern of permissible initial clusters. Although the second and third solutions are less elegant (and open to the objections raised above), it appears that either of them — or, most probably, some kind of combination, e.g. the gradual diffusion of underlying /d͡ʒ/ through the lexicon, to the detriment of an earlier /d-ʒ/ cluster — is more realistic. Hence, although with certain misgivings, the interpretation of /d͡ʒ/ as an independent phoneme can be accepted.

In sum, the question in the title can be answered as follows. The inventory of Hungarian phonemes includes three affricates: /t͡s/ as in *cica* 'kitten', /t͡ʃ/ as in *csúcs* 'peak', and /d͡ʒ/ as in *dzsem* 'jam'. Hungarian speech sounds further include three more affricates: [c͡ç] as one of the allophones of the voiceless palatal stop /c/ (*tyű* 'phew!'), [ɟ͡j] as one of the allophones of the voiced palatal stop /ɟ/ (*gyere* 'come!'), as well as [d͡z] as the coalesced (and then degeminated) realization of the cluster /d-z/ (*edzve* 'being trained'), as the voice-assimilated version of /t͡s/ (*kócból* 'out of hurds'), or as the result of the affrication of /z/, i.e. the insertion of [d] before it in casual speech (*pénz* [-nd͡z] 'money'). Just like any Hungarian consonant, these six speech sounds can also occur long (either as phonemic geminates or as coalesced clusters): [t͡s:] as in *moccan* 'budge', *vicc* 'joke' (/t͡st͡s/), *látszik* 'can be seen' (/t-s/); [t͡ʃ:] as in *loccsan* 'splash', *reccs* 'crack' (/t͡ʃt͡ʃ/), *kétség* 'doubt' (/t-ʃ/); [d͡ʒ:] as in *menedzser* 'manager', *bridzs* 'bridge' (/d͡ʒd͡ʒ/∼/d-ʒ/); [c͡ç:] as a variant of [c:] in *pottyan* 'plop', *fütty* 'whistle' (/cc/), *bátyja* 'his brother' (/cj/), *látja* 'he sees it' (/tj/); [ɟ͡j:] as a variant of [ɟ:] in *buggyan* 'spout up', *meggy* 'sour cherry' (/ɟɟ/), *hagyja* 'he allows it' (/ɟ-j/), *védje* 'let him defend it' (/dj/); and, finally, [d͡z:] as in *bodza* 'elder', *edz* 'train' (/d-z/): since /d͡z/ does not exist, geminate /d͡zd͡z/ is also impossible; hence, [d͡z:] can only arise through coalescence.

References

1. É. Kiss, K.—Papp, F.: A *dz* és a *dzs* státusához a mai magyar fonémarendszerben [On the status of *dz* and *dzs* in the phonemic system of present-day Hungarian]. *Általános Nyelvészeti Tanulmányok* 15. 1985, 160–172.
2. Fónagy, I.—Szende, T.: Zárhangok, réshangok, affrikáták hangszínképe [Spectrograms of stops, fricatives, and affricates]. *Nyelvtudományi Közlemények* 71. 1969, 281–344.
3. Kassai, I.: A magyar affrikátákról időtartamuk alapján [On Hungarian affricates: Their durational properties]. *Magyar Nyelvőr* 104. 1980, 232–245.
4. Kázmér, M.: *A magyar affrikátaszemlélet* [The interpretation of affricates in Hungarian]. Budapest, 1961.
5. Nádasdy, Á.—Siptár, P.: Issues in Hungarian phonology. *Acta Linguistica ASH* forthcoming.
6. Szende, T.: Magánhangzóközi affrikátáink természete [Intervocalic affricates in Hungarian]. *Magyar Nyelv* 71. 1975, 432–438.

FUNCTIONAL RANK ORDER OF CONSONANTS IN GEORGIAN

Eter SOSELIA

Language Typology Department
Institute of Oriental Studies
Academy of Sciences of Georgian SSR, Tbilisi, USSR

Phonological analysis of a language presumes not only establishing phonological units but also classifing them. There are two ways of classification: one - by distinctive features and the other - by distributional features. The latter is also known as the functional classification. Those two kinds of classification employed together make phonological analysis rather complete and finished. But the analysis would become even more perfect if distributional classes of phonemes could be ordered and in so doing their functional rank order obtained.

The model of Swede mathematician L. Garding (2) gives a possibility to get the order from any binary relation. The essence of the model is the following: A finite set $m = \{v_1, v_2 \ldots v_n\}$ is given and a binary relation (ω) is defined on it. As known relation ω can be presented as a subset of the set of pairs from m: $\omega \subset mxm$. x (a member of m) is in ω relation with y (a member of m), when $(x, y) \in \omega$. We can also write this in another way: $x\omega y$. The relation ω is order if and only if it is transitive (when for any x, y, z we have: if $x\omega y$ and $y\omega z$, then $x\omega z$), irreflexive (when there doesn't exist any x, so that x x), asymmetrical (when for any x and y we have: if $x\omega y$ and $y\omega x$, then x=y). We can get order from any binary relation in the following way: Firstly we define transitive closure of the relation ω ($\bar{\omega}$) - $x\bar{\omega}y$ if and only if there exists a sequence $s_1, s_2, \ldots, s_j$ of members from m, so that $x\omega s_1\omega s_2\omega \ldots \omega s_j\omega y$. The sequence $s_1, s_2, \ldots, s_j$ itself is called ω-chain. When $s_1 = s_j$ ω-chain is closed. As easily seen the relation $\bar{\omega}$ is transitive, but it may be not asymmetrical. The relation $\bar{\omega}$ isn't asymmetrical if and only if there exists at least one closed ω-chain. In that case we ought to define the relation δ: $x\delta y$ if and only if x and y are the members of the same ω-chain. The relation δ is that of equivalence and it devides m into classes that aren't empty and don't intersect. So we get the new set: $M = \{S_1, S_2, \ldots, S_k\}$, where $S_1, S_2, \ldots, S_k$ are the classes mentioned above. Let's define the relation Ω on M: $X\Omega Y$ (X and Y are members of M) if and only if there exist $x \in X$ and $y \in Y$, so that $x \in y$. The relation Ω is order.

Swede linguist B. Sigurd was the first to apply this model to a natural language (4). He modeled functional rank order of consonants in Swedish (established by distributional criteria). In order to define the primary binary relation he had to consider consonant clusters of monosyllables at the initial (anlaut) and at the final (auslaut) positions. Relation called "to have more

positional tendency to vowel" ($\dashv$) was regarded as primary binary relation and was defined like this: x$\dashv$y (where x and y are consonant phonemes) if and only if there exists yx-sequence in the initial clusters or -xy sequence in the final clusters. This relation, defined on the set of Swedish consonants, was irreflexive and almost asymmetrical. There were a few symmetrical pairs and as it was shown that they exist only in loanwords, these pairs were excluded from the primary relation. So the relation $\dashv$ became asymmetrical, but it wasn't order yet. For being order a relation ought to be transitive as well. The functional rank order of Swedish consonants was got by regarding transitive closure of the relation $\dashv$. It can be seen, that functional rank order gives more distributional information about a consonant system than any functional classification.

Using L. Garding's mathematical model in the way B. Sigurd did, it was possible to get the functional rank order of consonants in English (1).

We agree that L. Garding's method has quite successful results for the languages with simple consonant clusters (like Swedish or English). As for the languages with complex consonant clusters (like Georgian, where initial clusters may even consist of six consonants) the mathematical model ought to be used in a different way. In order to define primary binary relation we had to consider consonant clusters not only in monosyllables, but also in polysyllables, as consonant clusters of Georgian monosyllables present a very small part of the powerful set of Georgian consonant clusters.

The primary relation $\dashv$, defined on the set of Georgian consonants, has been neither asymmetrical nor transitive. Its transitive closure turned out to be the universal relation and the order was an empty set. This result gives minimum information about the character of Georgian consonant system. In order to get more information it is necessary:

1. to consider initial and final consonant clusters as separate systems;
2. to consider clusters of different length as subsystems of corresponding initial or final consonant cluster systems;
3. to modify the primary relation $\dashv$.

The primary binary relation could be defined separately for either of the systems in the following way:

(i) In the system of initial consonant clusters: x $\dashv$y if and only if there exists yx-sequence in the initial consonant clusters, where x isn't the last member of the clusters with three or four consonants (we don't consider clusters with two consonants).

(ii) In the system of final consonant clusters: x $\dashv$y if and only if there exists - xy sequence in the final consonant clusters, where x isn't the first member of the clusters with three or four consonants.

The relation, defined in a new way, isn't transitive for either of the systems, and so we need to consider the relation of equivalence (its note is $\sim$, corresponds to the relation $\bar{6}$ in L. Garding's model). The equivalence devides the set of

Georgian consonant phonemes into the classes, that aren't empty and don't intersect. The elements of the same distributional classes have equal positional tendency to vowel. Now we define relation $--\rightarrow$ on the set of these classes (it corresponds to the relation Ω in the L. Garding's model and is defined in the same way as Ω is). The relation $\rightarrow$ is order. It isn't the empty relation on either of the subsystems (a subsystem of initial clusters with three consonants is the only exception).

The functional rank order of Georgian consonant phonemes can be presented schematically (see: Figures 1, 2, 3, 4, 5, 6).

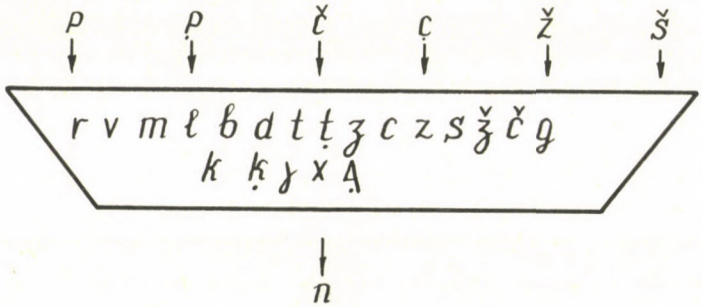


Figure 1. For the final clusters with three consonants

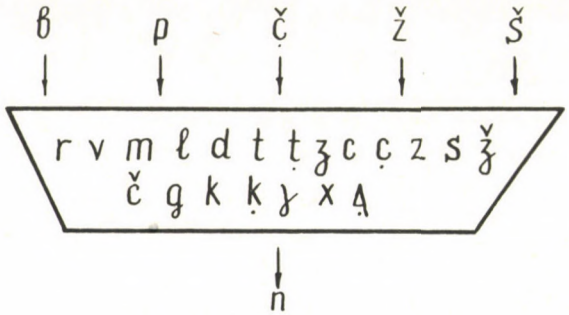


Figure 2. For the final clusters with four consonants

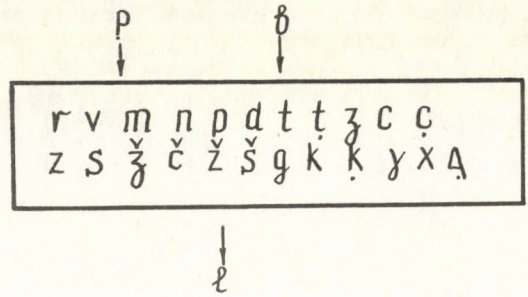


Figure 3. For the final clusters with five consonants

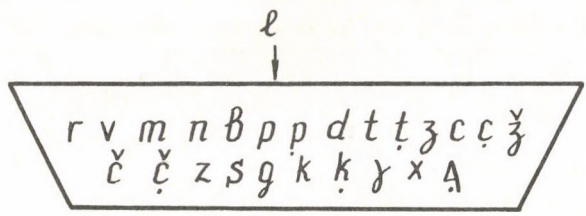


Figure 4. For the initial clusters with four consonants

It's important to note that the functional rank order of Georgian consonants doesn't confirm the supposition that the sonority of consonants is expected to grow in the position nearer to vowel (3).

Figure 5. For the initial clusters with five consonants

Figure 6. For the initial clusters with six consonants

References

1. DOXOPULO, M.: Syntagmatical Characteristics of English Consonant System. Dissertation (in Georgian). Tbilisi, 1971.
2. GÅRDING, L.: Relations and Orders. Studia Linguistica. 1955, N9.
3. MELIKISHVILI, I.: General Principle of Building Syllable and the Structure of Root in Proto-Kartvelian and Proto-Indo--European. Current Issues of the Academy of Sciences of Georgian SSR (in Georgian). Tbilisi, 1974, 4, p. 141.
4. SIGURD, B.: Rank Order of Consonants Established by Distributional Criteria. Studia Linguistica. 1955, N9.

PHONOLOGICAL REPRESENTATION AND 'GLOBAL PROGRAMMING'

Tamás SZENDE
Linguistics Institute, Hungarian Academy of Sciences, Budapest

It is the purpose of this paper, first, (i) to give a general account of (a polynomial word-level) 'phonological representation' considerably different from that as conceived of by post-SPE phonologies. Second, (ii) to argue that the multidimensional interrelationship existing between a low-level phonological representation and a (next-to-phonemic) phonetic representation may best be described in terms of what I call the Global Programming Principle. Finally, (iii) to show, and to exemplify on Hungarian, what kind of consequences the GP hypothesis might have on our conception of rules and rule domains.

(i) Word-level phonological representations are taken to be stratified abstract objects. One and the same representation may assume various forms at a deeper or shallower level of abstraction, depending on the actual perspective taken. Taking an arbitrary word form such as lássa 'see [ImplPSg]' and submitting it to analyses [= segmentations and structural parsings] of varying depth both in a syntactico--morphophonological and a phonemic perspective, we get different, but equally valid, results in terms of the set of primitives as well as their arrangement. The different results we get will be interrelated and derivable from one another. Such a stratified phonological representation of lássa 'see [ImplPSg]' will be approximately as follows:

(a) /la:ʃʃɔ/

(b) //#la:t#/] - /#j#/] - /#{ɔ, e}#/]

verb stem mood marker person marker

where (b) is supplemented by its historical antecedent (c) with which it is in a coordinate relation and at an identical level of abstraction:

(c) ///#lat | V, -front | # // = //#sV#/ -- /#C{a, e}#///

Representation (b) is related to (a) via morpheme structure rules and phonological rules; the same obtains for (c) and (b) with the proviso that this relation may involve non-productive morpheme structure rules and phonological rules (along with productive ones). Whereas (c) obviously has no role in speech production, i.e. it is 'extra-conscious' with respect to both speaker and listener, (b) is an active component of the speaker's mental processes at a 'pre-conscious' [vorbewusst] level, i.e. as a piece of unconscious knowledge that can be elevated to a conscious status, and, as such, it may acquire surface realization in special communicative situations (e.g. in spelling).

Level (a) of phonological representation--in the above example, /la:ʃʃɔ/ for lássa--must be invariant, i.e. discrete and of a constant form, whereas the corresponding word form in actual speech production is not. A set of interface rules must therefore be assumed to mediate between underlying forms and the ordered set of implementational instructions. In particular, I will assume that two types of interface rules, viz. 'smoothing' rules and 'gestalt' rules, will operate on phonological base forms.

'Smoothing' rules will effect transformations like /la:ʃʃɔ/ ➡ la:ʃ:ɔ

(where ➔ indicates level shift, and the omission of / is meant to reflect the fact that the form right of the arrow is neither phonological nor phonetic: rather, as a realizational program, it is an independent category constituting an intermediate level between those two). So we here will have to accept the assumption that morphemes are psychologically real, even if we cannot actually specify to what extent linguistic elements can be taken to be isomorphic with psychological facts (cf. e.g. Linell 1979, esp. 10--12). In the above example, smoothing rules turn a type (a) pre-implementational, intermediate phonological representation into the corresponding next-to-phonemic phonetic representation by removing the morpheme boundary feature from between /t/+/t/ and replacing /ʃʃ/ by ʃː via a pronunciation subroutine, in a way corresponding to the mechanism involved in the notion of Kiparsky's (1982) Bracket Erasure.

(ii) However, the form lássa --> laːʃːɔ will also undergo further operations, including the relativization of the [+long] component of /aː/. This follows from one of the gestalt rules--that of temporal organization--, a set of rules whose common property is that they involve a portion of an utterance as a whole. A correspondence like /aː/ ⟷ [aˑ] ~ [a], in fact, cannot be interpreted in terms of isolated segments if we wish to maintain the criterion of biuniqueness. The motivation for a derivation /aː/ --> [aˑ]~[a] can only be found in the structural effect of a word form as a whole, in the present case most immediately in the architecture -V:C:-, in particular, the occurrence of ʃː after aː, i.e. a (temporal) foot organization factor. The main properties of gestalt rules (omitting details) are as follows. (ii/a) Gestalt rules determine the utterance unit in speech production in a global way. This is unambiguously shown, in terms of my own experimental results, by 'sequence reduction' and 'sequence size truncation'. Another type of evidence comes from the stage of a child's first language acquisition where non-adult, 'incorrect' or 'crude' programming with respect to a given word form results in a disorderly arrangement of the articulatory components involved, one that does not match the order imposed by the phonological base form. For instance, Smith's (1973) data include squat surfacing as [ɡ̊ɔp], queen as [ɡ̊iːm], etc. by transposition of the bilabial component (cf. also Wilbur 1981, 411). (ii/b) The units undergoing gestalt rules may be of various sizes. They may involve single morphemes, but also several, semantically connected word forms (the latter case is observable primarily in sequence size truncation). (ii/c) In lenition processes, gestalt rules may exhibit varying effectiveness in modifying individual articulatory elements within a global articulatory program. For instance, of several units within a single word form, all of which are underlyingly specified as the same phoneme, e.g. /k/$_{\_\_3}$ in gyerekeknek 'children+Dat.', some will, and others will not, lose the element involved in the lenition process, in this case the stop component. This depends on the phonotactic position of the unit in question, the degree of lenition, the phonetic makeup of the segment, and so forth. In addition, it also depends on the feature/component itself; in vowel substitution errors, according to Shattuck-Hufnagel's (1986, esp. 124) data, the standard deviation of the feature [+tense] in erroneously substituted items exceeds the expected probability values several times more than that of [+back].

GP is made up by the totality of gestalt rules. The question of what sort of a cortical equivalent might be ascribed to GP is difficult to answer. Anyway, linguistic signs and processes are still considered to be best described, in terms of the functional hierarchy of the operation of language, by the model first proposed by Wernicke (1894). In essence, psycholinguistics also traditionally accepts this three-step mediation

model as one that confirms the authenticity of gestalt theories exactly "in the realm of perceptual organization" (see e.g. Osgood 1963, 146). In Wernicke's model, the levels wedged in between sensorium and cognitive representation are a bilaterally-connected "representation of specific 'gestalt' elements" and, on the speech production side, a "representation of motor commands (concepts of movements)" (cf. Creutzfeldt 1987, 5).

(iii) The domain of application of GP including gestalt rules is, obviously, phonetic implementation, especially that of lenition processes. In the rest of this paper my main concern will be the way gestalt rules fit into the rule hierarchy (P-rules, MP-rules, MS-rules, etc.) that is amply discussed in post-SPE phonologies (cf. e.g. Sommerstein 1977, Dressler 1985, Mohanan 1986). In terms of a typology of lenition processes observable in Hungarian, gestalt rules fit into this classical classificatory pattern rather badly. The facts are as follows. (iii/a) One particular lenition type, covering a set of essentially identical changes, may equally embody rules of diverse categories. 'Reduction', for instance, may simply be a change that we normally classify as a phonetic rule: the slight delabialization of a in változása 'its change' calls for that label. In other cases, reduction results in a change that can be characterized as a phonological rule that, by deleting a phonologically relevant feature, alters the phonological status (e.g., class membership) of a segment as in [m]→[w̃] (mondták 'say[Past3PPl]'). By eliminating a major classificatory feature, the realization may turn into the phonological base form of another lexemic alternant: by devoicing u in azután 'then' we get a result like az[v]tán which appears to be the 'fortis' version of aztán 'id.' (cf. Szende 1988, 182). (iii/b) It is not the case, however, that there is a complete and mutual overlap in that all types of lenition permit the occurrence of all possible rule categories. 'Truncation', for instance, is by definition a phonological category, not a phonetic one; indeed, there are clear examples (e.g. szóval 'in other words' --→ [so]) to show that truncated forms may fail to exhibit any further phonetic change (the omission of suffix being obviously not an instance of reduction). In other cases, it must be admitted, truncation and phonetic change may simultaneously occur within a single sequence, e.g. valami ilyesmi 'something like that' --→ [vɜ̃mijɛʃmɪ] where final i undergoes reduction by centralization and changes in height and degree of illabiality. (Note that in both cases we are faced with an independent phonetic rule applying or failing to apply at a different point in the same sequence, i.e. not one that involves the truncation site.) Consequently, the notions of truncation and phonetic rule are mutually exclusive. As for 'deletion' and 'loss', both lenition process types destroy a complete segment at the actual point in underlying form. The rules effecting these processes are undoubtedly of a non-phonetic character; but they may either be phonological like in cases of t-elision, e.g. in ezt 'this+Acc.', or result in morpholexemic switch as in the various versions of miért 'why' (cf. Szende 1988, 182). (iii/c) Scope properties are also non-relevant for the classification of lenition rules. Larger-scope processes, i.e. those involving a sequence of adjacent segments, can be realised by phonetic rules (cf. sequence reduction) as well as by morphophonemic or morpholexemic ones (as detailed above for cases of truncation). On the other hand, lenition phenomena involving single segments can also qualify as instances of any of these three rule types. (iii/d) Finally, it is appropriate to point out that rules responsible for lenition processes may also lead to results that do not lend themselves to a neat interpretation in terms of a linguistic system-oriented classification. Whenever sequence size truncation yields a realization that further undergoes elimination of backness contrast in a

vowel--as in ötkor --→ ötkör '5 o'clock' with [o]--→[ø]--the speaker in fact (over)applies vowel harmony in a way that, in terms of various lines of reasoning, can be taken to be of a phonetic, or morphophonemic, or (potentially) morpholexemic character.

The lack of correspondence between phonetic, morphophonemic and morpholexemic rules on the one hand and the set of gestalt rules on the other is conspicuous enough to make one wonder if those two systems of rules actually occupy different levels within the total system. However, the source of such mismatch is not that their structural descriptions reveal rule-governed phenomena of different depth: it is not the case that the former set of rules refer to phenomena restricted to underlying form and the latter account for events at some level intermediate between underlying and surface representation. (Aphasiacs' errors, in particular cases of syllable elision as in catholicize →→ /kaeθəlāyz/, solidification →→ /sālɔfəkēyšən/, demonstrate that syncope applies to underlying form, not (some level of) surface representation, cf. Schnitzer 1972, 24--29.). Rather, the difference actually lies in the fact that the rules categorized as above and gestalt rules can be stated for (a typologically diverse range of) allegro phenomena, whereas rules of the former type cover lento forms only. All that this distinction entails in itself, however, is that the number of gestalt rules is larger. But the punctum saliens of the comparison is that gestalt rules refer to sequences (utterance units) as wholes, whereas traditional types of rules refer to segments or concatenations of segments appearing between boundary features, even if their structural descriptions involve boundary features themselves as well. So, gestalt rules represent an independent category of rules; cover a set of phenomena exhibiting higher variability; and, consequently, phonetic/phonological/morpholexemic rules can, to a significant extent, be logically subordinated to them.

REFERENCES

1. CREUTZFELDT, O.: Inevitable deadlocks of the brain--mind discussion. GULYÁS, B. (ed.): The Brain--Mind Problem, Leuven 1987, 1--27. -
2. DRESSLER, W.: Morphonology. The dynamics of derivation. Ann Arbor, Michigan 1985. - 3. KIPARSKY, P.: Lexical morphology and phonology. Linguistics in the Morning Calm. Seoul 1982, 3--91. - 4. LINELL, P.: Psychological reality in phonology. Cambridge--London--New York--Melbourne 1979. - 5. MOHANAN, K.: The theory of Lexical Phonology. Dordrecht--Boston--Lancaster--Tokyo 1986. - 6. OSGOOD, C.: On understanding and creating sentences. American Psychologist XVIII, 1963, 735--751. - 7. SCHNITZER, M.: Generative Phonology--evidence from aphasia. University Park, Pennsylvania 1972. - 8. SHATTUCK-HUFNAGEL, S.: The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech. Phonology Yearbook III, 1986, 117--149. - 9. SMITH, N.: The acquisition of phonology: A case study. London--Cambridge 1973. - 10. SOMMERSTEIN, A.: Modern phonology. London 1977. - 11. SZENDE, T.: A note on morphophonological alternations in Hungarian. UAJb--Ural-Altaic Yearbook LX, 1988, 177--182. - 12. WERNICKE, C.: Grundriss der Psychiatrie. Leipzig 1894. - 13. WILBUR, R.: Theoretical phonology and child phonology: Argumentation and implication. GOYVAERTS, D. (ed.): Phonology in the 1980's. Ghent 1981, 403--429.

CONCRETENESS OF ABSTRACT PHONOLOGY? BIPHONEMATIC ANALYSIS OF THE FRENCH NASAL VOWELS

Domokos VÉKÁS
Scuola Normale Superiore, Pisa, Italy and
ELTE, Budapest, Hungary

Introduction

In contrast with the traditional position, the abstract generative analysis has adopted the view that the French nasal vowels (henceforth Ṽ) are derived from underlying sequences of oral vowel plus nasal consonant (e.g. Schane /9/). Within the same broad theoretical framework, but pretending a more direct correspondence to the physical details of articulation (Natural Generative Phonology), it was argued that such vowels are nasal (then monosegmental or monophonematic) underlyingly (e. g. Tranel /10/). I think instead that a closer look at the surface phonetic facts and at the phonetic behaviour of the "surface nasal vowels" can provide us arguments in favour of a biphonematic ("abstract") treatment.

Phonetically it should not be too strange to posit VN for a so-called nasal vowel: they show a lot of phonetic features in common:

(Partial) assimilation of nasality and e. g. labiality from a vowel is different: it is progressive (carry-over) and regressive at the same time in the case of the labiality, whereas it can only be progressive from Ṽ, and this is not surprising: even the beginning of such a vowel remains oral. On the other hand, a nasal consonant triggers progressive assimilation comparable to a Ṽ, so in French *dinde* and English *send* and Hungarian *rend* only part of the [d] remains oral. And a nasal also nasalizes (regressive assimilation) a preceding vowel (especially if tautosyllabic); in fact all three words above have nasalized vowels (to some extent). In analysing a Ṽ as a sequence one would avoid dealing with a strange kind of assimilatory phenomenon.

Van Reenen /11/ demonstrated that the increase in nasality (the proportional relationship between nose and mouth coupling) from the oral to the nasal part is relevant for the perception of a vowel as nasal. But phonologically obvious oral vowels, if followed by a nasal consonant, also show an increasing nasality. So Ṽ-s are destined for perceptual reasons to remain similar to VN sequences and cannot become totally nasal.

Ohala /7/ and Kawasaki /5/ conclude that shortening or weakening of a postvocalic nasal is in direct proportion with the percived nasality of the vowel. So the difference between VN and Ṽ phonetically can be insignificant.

The behaviour of nasal vowels on the time axis

As it is well known, especially after the seminal work done by Delattre and Monnot /4/, the French Ṽ-s are always longer than their oral "counterparts". At first glance the major duration may seem to be a simple correlate of nasality. So the distinctiveness of the feature "nasal" would be strengthend by another phonetic cue that could even outlive nasality, in these author's view.

But there is another possible interpretation: the exceptional length might be not a cue of a distinctive feature, but the sign of the bisegmental nature of a Ṽ.

In fact, there are some irregularities in the durational behaviour of Ṽ-s. Preceeding a so-called "consonne allongeante" their duration does not really increase as compared to the

position before a "consonne abrégeante". This behaviour seems to indicate that the nature of $\tilde{V}$-s is fundamentally different from that of oral vowels.

Before continuing to clarify the consequences of this irregularity, let's consider a parallel case in American English.

Vowel nasality is an important cue which contributes largely to the distinction of utterances as *bent* from *bet* (the nasal consonant being very short). However, vocalic nasality normally isn't considered to be a distinctive feature (as e. g. backness). Indeed, a "nasal vowel" is somewhat longer than an oral one (Malécot /6/, 223). Even if this greater duration is phonetically less prominent than the nasality, it nevertheless indicates a different prosodic behaviour.

Comparisons with [l] in the same positions suggest that the "practically insignificant vestigial nasal consonant" behaves as the lateral. In both cases, when [t] follows, the relation (interaction) between vowel and consonant is very intimate, [l] becoming vocalic to some extent, and the durational patterns are quite the same. For illustration here are some data from one male American speaker (each word uttered five times) in cs.:

	spend	spent	spelled	spelt
average duration (V + n/l + t/d)	31,5	22	36	24

In addition, final postvocalic [n] and [l] both show a considerably longer duration than in spent and spelt, respectively.

One could hardly consider such a $\tilde{V}$ monosegmental and the prosodically similar sequence vowel + lateral bisegmental (when followed by a voiceless consonant). It is interesting too that the nasal consonant is more attenuated when preceded by a somewhat longer vowel (*camp*) and not by a shorter one (*hint*)(Malécot /6/, 229.). We can then posit a nasal consonant who's exact duration is regolated also by some general compensatory phenomenon.

Experiment

I think the situation in French is similar. Measurements suggest that durations of oral vowel, $\tilde{V}$ and vowel + [r] ([r] being comparable to nasals in many aspects, and [l] having a different distribution) are quite the same befor [ʒ], the only lengthening (weak) consonant available in all these positions. On the other hand, befor the shortening [t] or [s] $\tilde{V}$ and Vr are very comparables, in evident contrast with oral vowels.

In order to attain high precision it appeared to be convenient to measure the whole duration from the beginning of the vowel to the release of the last consonant, the whole duration showing the same regularities.

The subjects were two Parisians between ages 20 and 30; the results were compared also to partial data (gathered for other purposes) from three other speakers of Northern France. The results for both subjects and all three $\tilde{V}$-s were consistent across tokens, those (deemed representative) presented here are from a female Parisian speaker (she does not distinguish *mettre* from *maitre* and pronounce *pate* et *patte* with minimal difference in durations) for the distinctions [ɛt - ɶ̃t - ɛrt, ɛʒ - ɶ̃ʒ - ɛrʒ]; the nonse words read five times all begun with [t]. a) = with final consonant, measured to the stop release or to the beginning of an embryonal schwa, with a precision of 1 cs; b) = without final consonant:

		[ɛt	ɶ̃t	ɛrt	ɛʒ	ɶ̃ʒ	ɛrʒ]
average duration	a)	34,5	42,5	41,5	43	41	42,5
(cs)	b)	14	26	28 (?)	28	28	(?)

And here are the data from a subject born in Bretagne (she did not make any difference between short and long vowels); real words come from a corpus of roughly 600 tokens (recordrd for general purposes):

	pote	fonte	porte	age	ange	marge
duration a)	38,3	45,7	44,8	49,1	47,7	53,2
(cs) b)	19	31	30	40	39	44 (?)

Measurements with two other subjects (both from Lorraine) show the same tendencies, but they do observe the difference of duration between long and short vowels, and the longs behave roughly as nasals do.

Discussion

Consonantal segments may have considerably different intrinsic durations, especially in final position, but we can assume that these differences are somewhat "neutralized" when another tautosyllabic consonant follows. We have seen that a Ṽ and a sequence Vr, but not an oral vowel, have a very similar timing behaviour when a consonant follows, and this phonetic fact is difficult to account for within a monophonematic analysis.

However, within a phonetically plausible bisegmental treatment of the Ṽ there arises the problem of the interpretation of sequences V + nasal traditionally considered as tautosyllabic, like in *bonne* or *caneton*.

According to Rialland /8/, in French some postvocalic consonants can sometimes constitute an independent syllabic nucleus, e. g. in *le bas r(e)trouvé* (which is different from *le bar trouvé*),*ça n(e) pouss(e) pas*, because such a consonant "ne produit pas sur la voyelle précédente les effets que l'on attendrait d'une consonne tautosyllabique" (p. 212). We can add to Rialland's arguments that a tautosyllabic nasal would nasalise much more a preceding vowel. One could account for this lack of nasalisation by assuming that if a language has a phonemic contrast for a feature, coarticulatory effects for that feature will be minimal. With Delattre's words: "Nasality being distinctive in French, it is essential that non-nasal vowels show no trace of nasality."

But, while a vowel as in *bonne* remains practically oral, that of *nez* is heavily nasalized (cf. Cohn /2/) by an evidently tautosyllabic nasal. This assymmetric resistence against assimilation seems to indicate a very loose contact between the vowel and the following nasal; and this loose contact is probably more or less the same for other consonants too, according to the similar difficulties involved in second language teaching: "Teaching American students of French [to pronounce final consonants] correctly, especially n, m and l, is not a small matter" (Delattre /3/, 113), and it is the same with Hungarian students.

In the case of a Ṽ as in *on* the contact between the two segments is closer, in words such as *bonne* it is looser than it would be with a "normal" coda position (as compared to some other languages). In the first case we should perhaps talk about diphthongs, and I think that every postvocalic nasal (and not only in the few words reported by Rialland) that does not trigger nasalization belongs to the second case. In the two cases the difference in the vowel-nasal contact is distinct enough to avoid confusion.

Conclusion

There are phonetic arguments (e. g. timing-behaviour) for positing VN for a Ṽ.

In the expanding variety of French without (consistent) length distinction between oral vowels, Delattre's prediction would have been more probable. However, the exceptional length of Ṽ-s is accompaigned by an exceptionally heavy nasalization despite Delattre's observations of "mediocre nasalization on spectrograms". In addition, I did not find significant (if any) durational difference between oral and nasal vowels before the lengthening [ʒ].

According to the current prevailing wave, phonology should attain concreteness /1/, but if concreteness is mistakenly confused with nearness to phonetic transcription (which imperfectly represents the complex phonetic reality) then even a so-called abstract solution, which most deviates from this imperfect representation might sometimes and somewhat paradoxically be more concrete (or real) than a "concrete" one, and the posited underlying forms might even correct some of the imperfections of the phonetic representation.

If there is a lesson to draw from the case of the French Ṽ-s it is that a truely concrete analysis should address the phonetic reality and not its (artificial) transcription based solely on impressionistic phonetic judgments.

References

1. BERTINETTO, P. M.: Phonetics, phonology, and the natural of it. In: Proceedings of the 11th I. C. Ph. S. Tallinn, 1987.
2. COHN, A. C.: Phonetic Rules of Nasalization in French. UCLA Working Papers in Phonetics 69. 1988, 60--67.
3. DELATTRE, P. Comparing the Phonetic Features of English, French, German and Spanish. Heidelberg, 1965.
4. DELATTRE, P.--MONNOT, M.: The Role of Duration in the Identification of French Nasal Vowels. IRAL VI/3. 1968, 267--88.
5. KAWASAKI, H.: Phonetic Explanation for Phonological Universals: The Case of Distinctive Vowel Nasalization. In: Ohala, J. J.--Jaeger, J. J. (eds.): Experimental Phonology. Academic Press, 1986, 81--103.
6. MALECOT, A.: Vowel Nasality as a Distinctive Feature in American English. Language 36. 1960, 222--29.
7. OHALA, J. J.: Phonetic explanations for nasal sound patterns. In: Ferguson, C. A.--Hyman, L. M.--Ohala, J. J. (eds.): Nasálfest. Papers from a Symposium on Nasals and Nasalization. Stanford, 1975, 289--316.
8. RIALLAND, A.: Schwa et Syllabes en Français. Wetzels, L.--Sezer, E. (eds.): Studies in compensatory lengthening. Dordrecht, 1986, 187-226.
9. SCHANE, S.: French phonology and morphology. MIT. 1968.
10. TRANEL, B.: Concreteness in Generative Phonology: Evidence from French. Berkeley, U.C.P. 1981.
11. VAN REENEN, P.--VAN DEN BERG, H.: A problem of assimilation between nasal vowel and preceding nasal consonant. In: Proceedings of the 11th I. C. Ph. S. Tallinn, 1987.

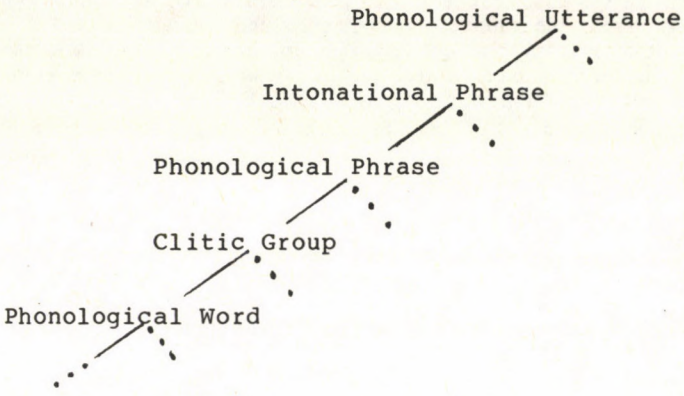
The role of the clitic group in prosodic phonology

Irene VOGEL
University of Delaware

1. Introduction

One current model of the organization of the postlexical phonology can be represented as in (1).

(1)



This is the hierarchical arrangement of prosodic constituents found, for example, in Nespor and Vogel (1986) (henceforth N&V). It differs in several ways from Selkirk's (1978) original model of prosodic phonology, one of which is the presence of a constituent between the Phonological Word (PW) and the Phonological Phrase (PP). This constituent, the Clitic Group (CG), is probably the most controversial of the prosodic constituents, and it is the one that will be the focus of the present paper. In particular, I will first briefly consider whether it is indeed needed as a prosodic constituent. I will conclude that it is, and then show that the definition of the constituent found in N&V is inadequate in a particular area, specifically where compounds are concerned. Finally, I will propose a modification of the original definition that allows us also to handle compounds.

2. Do we need the Glitic Group?

In general, a string is considered to be a constituent of grammar if (a) there are rules that have precisely that string as their domain of application or (b) there are rules that need to refer to it in some other way in their formulation. As far as the CG is concerned, it was first proposed by Hayes (1984) as the domain of two rules of English, one of which is a fast speech rule, and several metrical constraints observed in a number of English poems. The only rule of "normal" English - the type we

would require a grammar to handle - is the Palatalization rule in (2).

(2) s, z --> š, ž / __ š, ž (Hayes 1984)

This rule operates in (3a) changing [z] to [ž] within a CG, but not in (3b), across CGs.

(3) a. [Is Sharon] [coming?]
CG CG
ž

b. [Clara's] [shadow]
CG CG
*ž

Actually, Hayes mentions that the rule may even apply across CGs in fast or sloppy speech. Thus, the implications of Hayes' analysis for "normal" English, and more generally for a phonological constituent between the Phonological Word and the Phonological Phrase is not overwhelming. Fortunately, however, there are other rules in a variety of languages that apply within the same domain, as shown in N&V.

Why then should there be any doubt about the existence of the CG as a constituent of phonology? One reason is that the CG is often isomorphic to the PW, as can be seen in the Italian and equivalent English sentences in (4), and we might wonder whether both are, in fact, needed.

(4)

CG	CG	CG	CG	CG	CG	CG
PW	PW	PW	PW	PW	PW	PW
Carla	balla	bene	soltanto	quando	Franco	canta
Carla	dances	well	only	when	Frank	sings

It is suggested in Nespor and Vogel (in press), furthermore, that, at least in some languages, the CG is not relevant in constructing the grid for rhythmic phenomena because of its frequent isomorphism to the PW and because one of the characteristics of clitics is that they do not themselves bear stress. What I would like to suggest here, however, is that the situation may actually be just the opposite. That is, instead of the CG <u>not</u> being relevant for rhythmic phenomena because of the inherent stressless nature of clitics, it may be that it is relevant for this very reason. Since clitics do not bear stress, they have to be grouped into units with other items that do, and this would appear to be precisely what is being called the Clitic Group.

In French, for example, as can be seen in (5), stress is placed on the last syllable of a CG, even when this is a clitic.

(5) a. changez-lé 'change it'

b. allez-vous én 'go away'

Clitics also participate in stress assignment in Latin, as can be seen in (6).

(6) a. fémina 'the woman (nom)'

feminăque 'and the woman'

*femínăque

b. úndĭque 'everywhere'

undĭque 'and from there'

While the familiar Latin Stress Rule would place stress on the antepenultimate syllable in the first form in (6a), if a clitic such as -que is present, a different rule applies, assigning stress to the syllable directly preceding the clitic, as in the second form. The first rule would incorrectly place stress on the antepenult again, as shown in the third form in (6a). In (6b), furthermore, we see that the application of the first rule yileds a form with one meaning while the application of the other rule yields a form with a different meaning. Thus, we need two different domains for stress assignment in Latin, one of which must include clitics.

If it turns out that the CG is particularly relevant for stress related phenomena, this would constitute even stronger evidence for its role as a constituent in phonology. That is, if we can eventually predict, at least in the majority of cases, the domain of phonological rules on the basis of the type of operation they involve, this will constitute a fairly stong type of argument for the existence of the constituent in question. Along these lines, we might expect that the CG would not only be relevant for stress assigning rules such as those in French and Latin, but also for other types of rules that depend on stress. In fact, Maiden (1988) discusses a number of CG domain rules in a variety of Italian dialects that seem to support such a hypothesis. That is, a number of the rules involve vowel reduction and syncope in relation to the position of the stressed syllable - in the CG. This is illustrated in (7) with examples from a dialect spoken in southeastern Italy, Montefalcone.

(7) Montefalcone (Maiden 1988)

a. /sábatu/	[sábbətə]	'Saturday'
b. /fatíka/	[fatíjə]	'toil'
c. /pappagállo/	[pappajállə]	'parrot'
d. /la kánta/	[a kándə]	'he sings it'

e. /sóra + ma/ [sórəmə] 'my sister'

f. /kognáta + ma/ [kajnátəmə] 'my sister-in-law'

In these forms we see that /a/ --> [ə] to the right of the stressed vowel within the domain of the CG. A stressed /a/ and any unstressed ones to the left of the syllable bearing primary stress remain [a].

On the basis of the phenomena considered thus far, it seems that there is fairly strong evidence that there is indeed need for a constituent in phonology that includes clitics, along the lines of Hayes' original proposal, and that subsequently developed elsewhere. Let us now turn to the way in which the CG constituent is defined.

3. Defining the Clitic Group

Before actually examining the definition of the CG, I will first motivate the need for a constituent between the Phonological Word and the Phonological Phrase in Hungarian, since this is the language that will serve as the basis for the rest of the paper. In (8) - (11), we see that the domain of Vowel Harmony is the PW, where this constituent consists of a stem plus any affixes to its right.

(8) a. kert - em - ben 'in my garden'
garden my in

b. udvar - am - ban 'in my yard'
yard my in

(9) a. fel - bukkan 'appear suddenly'
up appear

b. oda - néz 'look there'
there look

(10) a. nyak - kendő 'neck tie'
neck tie

b. fürdő - szoba 'bathroom'
bath room

(11) a. kerék - pár - ok 'bicycles'
wheel pair pl

b. hang - verseny - ek - re 'to the concerts'
sound competition pl to

The items in (8) show that harmony takes place in a string that includes a stem and all following suffixes, that is, throughout a PW. In (9), however, we see that harmony does not apply between a verb and a preverbal element to its left. Similarly, it is

blocked between the members of a compound, as shown in (10). When suffixes are attached to a compound as in (11), however, harmony does take place in the string that groups these suffixes with the last member of the compound, indicating that such as string is a PW, despite the fact that this is not a constituent in the morphological structure of the word. That is, the suffixes pertain to the entire compound, not just the final member (cf. N&V).

Let us now consider another aspect of Hungarian phonology: stress placement. As is well known, word stress always falls on the first syllable, as shown in (12), where the underlined vowel is the one with primary stress.

(12) a. Amerika 'America'

b. amerikai 'American'

c. amerikaiak 'Americans'

d. amerikaiakat 'Americans + acc'

In compounds, too, the primary stress falls on the leftmost syllable, as can be seen in (13). The primary stresses of the other members of the compound are reduced. An analogous pattern can be observed in verbs with a preverbal element, where the main stress falls on the preverbal element, as in (14).

(13) a. fürdő - szoba --> fürdőszoba
bath room 'bathroom'

b. nyak - kendő --> nyakkendő
neck tie 'necktie'

(14) a. oda küldenek --> oda küldenek
there send Pe3pl 'they dispatch'

b. kenyeret eszik --> kenyeret eszik
bread eat Pe3pl 'they eat bread'

When clitics are present, however, the primary stress in no longer necessarily on the leftmost syllable, but rather it is on the first syllable of the lexical item, as illustrated in (15) and (16).

(15) a. az egyetem 'the university'
the university

b. és ha írta 'and if he wrote'
and if wrote Pe3sg

(16) a. Gábor is 'Gabor too'
Gabor too

b. Péter meg 'and Peter'
Peter and

In phrases, on the other hand, each lexical item has its own stress under most circumstances, as in (17). (See Vogel and Kenesei (1987) for a discussion of cases in which phrasal stress patterns may be somewhat modified.)

(17) a. sárga virágok 'yellow flowers'
yellow flowers

b. keservesen sír 'he cries bitterly'
bitterly cries Pe3sg

What the different domains of Vowel Harmony and stress assignment show is that we need two different constituents in order to account for these phenomena, particularly where compounds are concerned. While Vowel Harmony does not apply within a domain that includes the two members of a compound, stress assignment does. Since the domain for Vowel Harmony is clearly the PW, we need a larger domain for stress assignment. This cannot be the Phonological Phrase, however, since in phrases, each lexical item has its own word stress. Thus, what is needed is a phonological constituent between the PW and the PP, the obvious candidate for this being the Clitic Group. Furthermore, it is interesting in light of the suggestion made above, that the phenomenon that seems to require the existence of the CG in Hungarian is precisely a stress assignment rule.

Let us now turn to the matter of defining the CG. The definition provided in N&V is given in (18).

(18) CLITIC GROUP FORMATION (Nespor and Vogel, 1986:154-155)

i) Clitic Group Domain

The domain of CG consists of a PW containing an independent (i.e. nonclitic) word plus any adjacent PWs containing
a) a DCL, or
b) a CL such that there is no possible host with which it shares more category memberships.
(DCL = Directional Clitic; CL = Clitic)

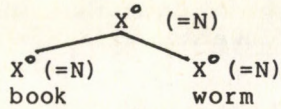
ii) Clitic Group Construction

Join into an n-ary branching CG all PWs encluded in a string delimited by the definition of the domain of CG.

Essentially, (18) defines the CG as a string consisting of a PW plus adjacent clitics of various sorts. What was seen above in relation to Hungarian is that the members of a compound each form their own PW for the purposes of Vowel Harmony. According to (18), we would thus also expect the members of a compound to form separate CGs. As we have seen on the basis of stress assignment,

however, this is not correct. What seems to be the crux of the problem is what level of a compound is relevant for CG formation. That is, if we look as the structure of a compound, we typically find something like the representation in (19), where we have one X dominating two other X s.

(19)



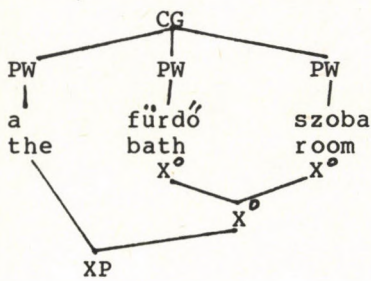
The question that arises at this point is whether it is the uppermost X° or the lower ones that are relevant for CG formation. In the former case, the entire compound would be in one CG, while in the latter case, each member of the compound would be in a separate CG. What I propose is that both possibilities exist and that they represent two settings of a parameter that a language can choose from regarding CG formation. Such a parameter can be incorporated into the CG formation rule to give the revised version seen in (20) (cf. Vogel 1988).

(20) Clitic Group Domain (revised)

The domain of CG consists of a PW or PWs containing any independent word(s) dominated by the $\left\{\begin{matrix}\text{highest}\\\text{lowest}\end{matrix}\right\}$ X° node plus any adjacent PWs containing...

Hungarian chooses the setting of "highest X°" node, as illustrated in (21).

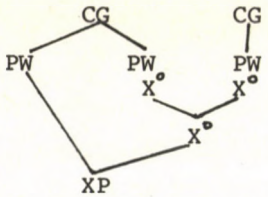
(21)



'the bathroom'

In a language that chooses the other option, the "lowest X°", we would have the CG and morphological structures shown in (22). Such a language appears to be Taiwanese (Chiang, 1988).

(22)



Finally, it should be pointed out that the proposal made here that allows us to treat compounds as single CGs is not an ad hoc solution valid only for the problem posed by Hungarian. Instead, the same revision of the CG formation rule appears to be necessary in other, unrelated, languages as well. One such language is Swedish (L. Bailey, personal communication). While phonotactic constraints indicate that the two members of each compound in (23) must be separate PWs, their patterns of pitch accent indicate that they must nevertheless constitute single constituents at some level.

(23) a. präst - arm 'priest sleeve'

b. svensk - språkig 'Swedish speaking'

In order for (23a) to constitute a single PW, the st would have to be syllabified with the following syllable, which it is not here. On the other hand, (23b) could not form a single PW under any circumstances, since nskspr is not a possible medial consonant sequence. Both compounds, however, only bear one pitch accent, which indicates that at least for this phenomenon they must constitute a single domain. As in Hungarian, this cannot be the Phonological Phrase since each lexical item in a phrase has its own pitch accent. Thus, the conclusion we are led to is that the domain of pitch accent assignment is the CG (cf. Bailey, in progress), and that Swedish, too, chooses the "highest X " in compounds for the purposes of CG formation. This result, furthermore, lends additional support to the hypothesis advanced above that the CG may be the constituent that is particularly relevant for stress and stress-related phenomena.

4. Conclusions

What has been demonstrated above is, first of all, that we do need a constituent in the prosodic hierarchy between the Phonological Word and the Phonological Phrase. Secondly, it was shown that the definition of the Clitic Group given in N&V is inadequate as far as compounds are concerned. In order to correct this problem, a parameter was proposed that allows languages to treat the members of a compound either as separate CGs or as a single one.

REFERENCES

Bailey, L. (in progress) The Phonology of Swedish Pitch Accent. University of Delaware: Ph.D. Diss.

Chiang, W.-Y. (1988) The phonological word in Taiwanese. Paper read at the 63rd Annual LSA Meeting. New Orleans.

Hayes, B. (1984/ in press) The prosodic hierarchy in meter. In P. Kiparsky and G. Youmans (eds.) Perspectives on Meter. New York: Academic Press.

Maiden, M. (1988) On the internal prosodic structure of prosodic constituents. Ms. University of Bath.

Nespor, M. and I. Vogel (1986) Prosodic Phonology. Dordrecht: Foris.

Nespor, M. and I. Vogel (in press) On clashes and lapses. Phonology. 6.

Selkirk, E.O. (1978/1981) On prosodic structure and its relation to syntactic structure. In T. Fretheim (ed.) Nordic Prosody. II. Trondheim: TAPIR.

Vogel, I. (1988) Prosodic constituents in Hungarian. In P.M. Bertinetto and M. Loporcaro (eds.) Certamen Phonologicum. Torino: Rosenberg & Sellier. 231-250.

Vogel, I. and I. Kenesei (1987) The interface between phonology and the other components of grammar. Phonology. 4. 243-263.

SPEECH FEATURE PERCEPTION BY PATIENTS USING A SINGLE-CHANNEL VIENNA 3M EXTRA-COCHLEAR IMPLANT

Eva AGELFORS and Arne RISBERG
Dept. of Speech Communication and Music Acoustics, Royal Institute of Technology (KTH), Stockholm, Sweden

The speech perception ability, reported from cochlear implant patients both with single and multiple channel implants, varies widely. In a single-channel cochlear implant, the electrode is either placed in the middle ear close to the round window, or inserted a few millimeters into the cochlea. In a single-channel implant, it seems that mainly time-intensity information in a speech signal can be transmitted. The good speech understanding reported from subjects using single-channel devices seems, however, to indicate that they also have some possibility to identify vowels and some consonant features.

In the Swedish cochlear implant project, ten post-lingually deaf patients have been implanted with a single-channel implant with extra-cochlear electrode. After implantation, the patients went through a longer, structured training and test program. Testing was made 1, 3, 6, 12, 24, and 36 months after surgery. In the test battery, measurements of frequency, time discrimination, and speech perception ability with and without simultaneous lipreading were included.

The results of tests in the perception of speech features obtained from ten implanted patients, 12 months after surgery, showed that almost all of them are able to perceive some prosodic information. Only one of the subjects has ability to identify the Swedish vowels with any accuracy. She has also some speech perception ability without support of lipreading. Consonant perception is mostly based on temporal distinctions. The results of the consonant tests show that the subjects make use of informations regarding amplitude and voicing, and presence and absence of friction.

Introduction

It has long been known that stimulation of the auditory nerve with a weak electric current results in auditory sensation. As early as in 1790, Volta made experiments with electrical stimulation of his ear. During the last two decades the research in the fields of electronics, audiology, speech science, and surgery has made it possible to introduce a limited world of sound to profoundly deaf patients. This has been carried out by cochlear implants which electrically stimulate the auditory nerve. Since the beginning of the seventies, House at the Ear Research Institute in Los Angeles has been implanting deaf patients with a simple single channel cochlear implant (4). At the same time, research and development have been going on at several laboratories both on single-channel and multi-channel devices. In the eighties, the number of patients with implants has been growing rapidly. It has been estimated that today over 4000 patients have received cochlear implants of various types. As more advanced systems are introduced, the number of patients with implants will undoubtedly grow fast.

In a single-channel extra-cochlear implant, the active electrode can be placed in the round window niche or inserted in the bone of the promotory. This placement of the active electrode results in that a large number of neurons is stimulated simultaneously. During the last year, very good speech understanding without support of lipreading has been reported from subjects using this type of implant. (3)

In the Swedish cochlear implant project reported here, a single-channel implant developed in Vienna and manufactured by 3M is used (6). The project is run at the Department of Audiology of the South Hospital (Södersjukhuset), Stockholm, in cooperation with our department.

Subjects

Ten post-lingually deaf patients participated in this study. Criteria for operation were total acquired deafness, an active cochlear nerve, no benefit from hearing aids, strong motivation, a good social back-up, and auditory memories, which excluded them who were born deaf.

Some data on the ten patients. Table 1

Subjects:	*S1*	*S2*	*S3*	*S4*	*S5*	*S6*	*S7*	*S8*	*S9*	*S10*
Sex:	*F*	*M*	*M*	*M*	*F*	*F*	*M*	*M*	*F*	*M*
Age:	*29*	*54*	*52*	*42*	*54*	*25*	*49*	*22*	*55*	*50*
Years of deafness:	*4*	*27*	*7*	*20*	*46*	*14*	*10*	*2*	*1*	*6*

Progressive hearing-loss: S1, S6, S7, S9, S10 .
Meningitis: S2, S5, S8
Ototoxic drugs: S4 and Scull fracture: S3

Material and methods

After implantation, the patients go through a longer, structured training and test program. Testing is made 1, 3, 6, 12, and 24 months after surgery. In the test battery, measurements of frequency and time discrimination and speech-perception ability are included. The test battery is developed at our department and primarily used for assessing communication competence of severely hard of hearing persons with a view to determining appropriate rehabilitation strategies (5). The test program is based on a model of speech perception that includes a number of processing levels. At each level, tests are used to determine how well the speech perception system of the patients is functioning. In this model, the two lowest levels are: **I** signal transformation and **II** signal analysis. At higher levels, **III** phonetic interpretation, the ability to extract basic linguistic information, is tested and on the highest levels, **IV** information processing ability and **V** linguistic interpretation. In this study, results on levels **II, III,** and **IV** have been evaluated 12 months after implantation .

The following signal analysis measurements (level **II**) were used:

1. Frequency discrimination with sinusoidal signals.
   Measurements are made in the frequency range 125Hz-3000Hz. This test is seen as a general test of the signal analyzing capacity.

2. Frequency discrimination with a band-pass filtered pulse-train.
   Band-pass filtered pulse trains of white noise with a band width of 1000 Hz and mid-frequency 500, 1000, 2800 Hz. The stimulus had either a constant repetition rate, or it was frequency modulated at 2 Hz. Testing was done with pulse repetition rates of 125 and 250 c/s. The test was meant to simulate intonation of a male and female speaker.

3. Gap-detection with band-pass filtered white noise.
   The test measures time-resolution by means of a short interruption in a two seconds long band-pass filtered noise signal. Bandwidth 1000 Hz, mid-frequency 500, 1000, 2800 Hz

4. Periodic/non periodic signals.
   The test measures identification time for periodic and non periodic signals. In this test, two signals are used, a pulse train with the repetition frequency 120 Hz and white noise. The signals are band-pass filtered with the same filters as used in test 2 and 3.

Speech tests

The perception of speech is closely related to the detection of both phonetic and prosodic structures of the speech signal. The prosodic information does not change the time domain as fast as the phonetic information. It contains the information of F0, intensity and temporal spacing of gross acoustic events. It is generally assumed that speech perception with a single-channel implant is confined mainly to the prosodic information.

The test of speech perception (level **III, IV**) consists of rhyme tests based on acoustic differences as identification of phonemes, syllables, and word stress, intonation, spondee words, and words in context. Word-lists with two or three response alternatives are used in the rhyme tests. The test list ranges in difficulty from gross discrimination to minimal phonetic contrasts. In the test battery used, a test with 12 known spondee words is also included.

The test equipment was a computerized self-instructed test system. Each patient was tested individually and coupled to the test system over the line input of the implant and they adjusted the controls on his/her own stimulator unit to a comfortable level. The patients could repeat the stimulus as many time they wanted before answering.

Results

The results of feature perception in the rhyme tests expressed as total range and median values by ten cochlear implant patients is shown in Fig.1. Fig. 2 shows the results from the test with 12 known spondee words. The patients S5 and S6 found the test too difficult. Fig. 3 shows the relation between the frequency discrimination ability at 125 Hz for sinusoidal tones and the result on the spondee test.

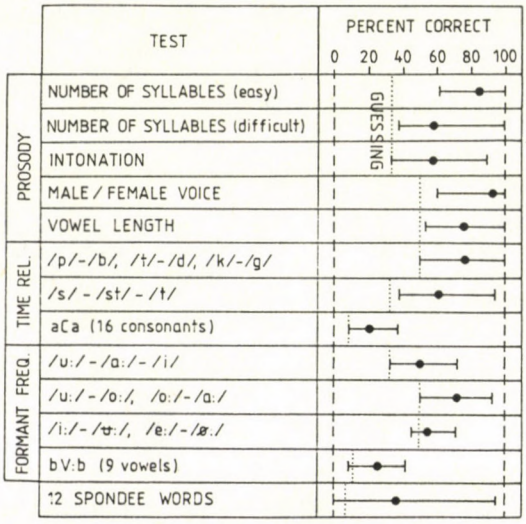


Fig. 1. Results on the speech feature test and the test with 12 known spondee words for the ten subjects at the testing 12 months after implantation. Total range and mean values are shown.

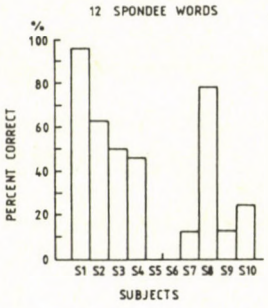


Fig. 2. Results on the test with 12 known spondee words.

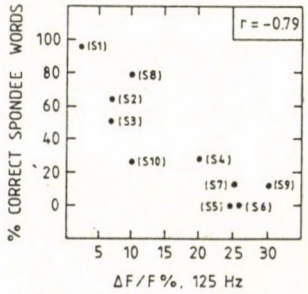


Fig. 3. Relationship between percentage correct recognition of spondee words and frequency discrimination ability at 125 Hz.

Discussion

The figures show large variation in results on the different tests. This is typical for cochlear implant patients and has been reported for both single and multi-channel implants. Better results are on the average obtained on the prosody test. The good correlation between frequency discrimination ability at 125 Hz (Fig. 3) and percent correct on the spondee test indicate that a single-channel implant mainly transmits low-frequency information. The correlation obtained (r=-.79) is one of the highest between any of the signal analysis measures and the ability of word identification . Two subjects stand out as "star"-patients, S1 and S8, see Figs. 2 and 3. Detailed studies have been made of S1:s ability to identify vowels. (1) These studies indicate that she has reasonably good ability to perceive both the first and the second formant of the vowels. On the test with nine long Swedish vowels, she scored 67% correct four years after implantation. These results contradict the hypotheses that a subject with a single-channel implant can only get low-frequency time-intensity. Some subject apparently get more information in the speech signal. The mechanism behind these differences in result is at present not clear. An explanation that has been suggested is that in some patients', remaining hair cells are stimulated. (2).

References

1. AGELFORS, E.--RISBERG, A. The identification of synthetic vowels by patients using a single-channel cochlear implant, Speech Transmission Laboratory, Quarterly Progress and Status Report (KTH, Stockholm) 2-3. 1987, 31--38.

2. BROKX, J.P.L.--HOMBERGEN, G.--CONINX, F. Relations between audiometrical thresholds of potential cochlear implant patients and their performance in preoperative psycho-physical tests with electrical stimulation. Scandinavian.Audiology 17. 1988, 217--222.

3. HOCHMAIR-DESOYER, I.J.--HOCHMAIR, E.S.--BURIAN, K.--FISCHER, R.E.: Four years of experience with cochlear prostheses. Medical Progress Through Technology 8. 1981, 107--119, 1981.

4. HOUSE, W.F.: Cochlear implants. Annals of Otology, Rhinology and Laryngology, Suppl. 27, 85, No. 3-2. 1983.

5. RISBERG, A.--AGELFORS, E. An audiological test battery for speech communication diagnosis in severe hearing loss. In T. Lundborg, ed. Diagnostic and Therapoeutic Problems in Postlingual Auditory Communication Handicap. Scandinavian Audiology, Suppl. 18. 1983, 5--10.

6. v. WALLENBERG, E.L.--HOCHMAIR-DESOYER, I.J.--HOCHMAIR, E.S.: Speech processing for cochlear implants. Proceedings of the Seventh Annual Conference of the IEEE/Engineering in Medicine and Biology Society, Chicago, 1985. 1114--1118.

ACOUSTIC-PERCEPTUAL STUDY OF CV SEQUENCES PRODUCED BY TWO GLOSSECTOMIZED SPEAKERS

Ann-Marie ALMÉ and Olle ENGSTRAND
Institute of Linguistics
University of Stockholm, Stockholm, Sweden

Introduction

Glossectomy is the surgical removal of all or part of the tongue, usually performed to treat carcinoma of the tongue. If larger parts of the tongue are missing, oral vegetative functions are severely impaired. From a phonetic point of view, a variety of symptoms arise which may interfere with speech production and intelligibility of speech.

The general aim of our current research project "Speech after glossectomy" is to use the methods of experimental phonetics to create an empirically solid knowledge base concerning the effects of various kinds and degrees of glossectomy on speech production and understanding. Special attention is paid to compensatory articulation in relation to the type and extent of tongue resection.

In this paper we report quantitative data bearing on the production and perception of vowels and consonants in the speech of two glossectomized Swedish subjects.

Methods

The subjects were one male (MGS) and one female speaker (FGS). Subject MGS had undergone subtotal glossectomy and partial neck dissection 6 years prior to the recording. The female subject had undergone total glossectomy, radical neck dissection and partial mandibulectomy one year prior to the recording. Data were also obtained for one normal male speaker (NS).

The speech sample consisted of a word list and three short text passages. The word list was made up of 51 words with the structure /CV:1/. C stands for all Swedish consonant phonemes which are possible in morpheme-initial position; and V stands for one of the tense (long) vowels /i:/, /a:/ and /u:/; /a:/ is a back, slightly rounded vowel. These vowels approximate three of the extreme cardinal vowel points.

For the acoustic analysis, broad-band spectrograms were made of all CVC-words and lexically stressed vowels /i:/, /a:/ and /u:/ in the text material. The first and second formants (F_1 and F_2) were measured at the first point during the voiced segment where F_2 had attained its maximum or minimum value.

In the perceptual study, 14 undergraduate speech pathology students with normal hearing listened twice to the randomized CVC-syllables. Their task was to identify the first consonant and

the vowel. Half of the listener panel was asked to identify the consonants in the first round and the vowels in the second round, and vice versa for the other half of the panel. The listeners were instructed to give an answer to each item and to guess when uncertain.

Results

Mean frequency values (in Hz) for the first two vowel formants are illustrated in fig. 1. The figure, displaying the data in the F_1 - F_2 plane, shows differences as well as similarities between speakers. Both glossectomees distinguish formant frequencies for /i/, /a/ and /u/, but the differences are generally smaller compared to the normal speaker. For all speakers, vowels in connected speech are reduced compared to vowels in words in isolation. MGS reduces /a/ and /u/ relatively more than NS.

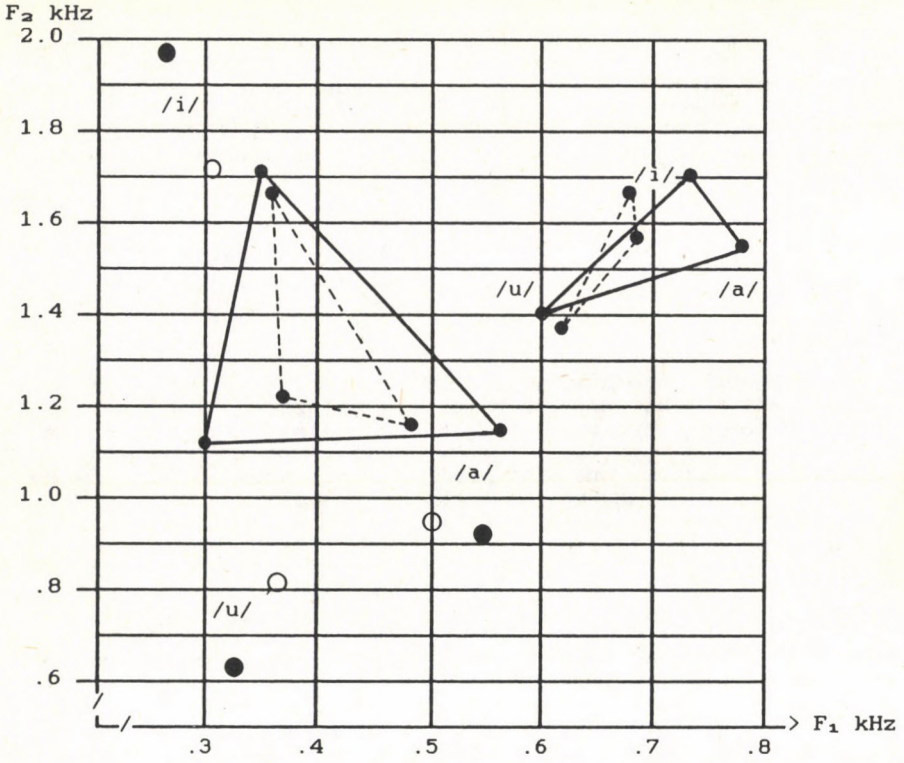


Figure 1. F_1 vs. F_2 plot for the tense vowels /i a u/ produced by the male subject MGS (left) and the female subject FGS (right). Mean values from word list (solid line) and connected speech (dotted line). Values for the normal speaker NS are indicated by filled and unfilled circles for word list and connected speech, respectively.

There are, however, considerable differences between the respective F_1 and F_2 effects. The glossectomees display a more restricted range in F_2 than in F_1, especially the female subject. This effect is expected since F_1 variation reflects the degree of jaw opening (1), and this articulatory parameter should not be greatly affected by glossectomy. F_2, on the other hand, is known to relate to the anterior-posterior movement of the tongue in normal speech. In consequence, this dimension is largely out of control in the speech of the glossectomees. This result is also in agreement with the data published by Morrish (2).

The glossectomized subjects produce both labial and lingual consonants with the lips as the primary articulator. Acoustic measurements reveal considerably less differentiation between consonants than the normal speaker. Formant transitions are much less dynamic than in normal speech.

Results from an identification test using consonants produced by the male subject MGS are shown in table 1. Confusions were analyzed in terms of manner and place of articulation, voiced versus voiceless, and misidentifications within or out of class (e.g. if a plosive was misidentified as another plosive it was regarded as a confusion in class; if identified as a fricative it was a confusion out of class). In terms of manner of articulation, most confusions are within class except for the liquids. The lateral was mostly heard as /v/.

The majority of erroneous identifications were in the place feature. As expected, bilabials and labiodentals were well identified with only few errors out of class. For the other consonants, however, there was a high percentage of errors out of class except for the laryngeal /h/. Non-labial sounds were often perceived as labials, but to a lesser extent than expected in view of the subject's consistent labial substitution for all target linguals. For example, target /p/ was identified correctly in 93% of cases, whereas target /t/ and /k/ were heard as /p/ in only 43% and 50% of cases, respectively. Apparently, MGS compensates for the missing tongue articulation. To identify the acoustic and articulatory basis of this compensatory adjustment, direct articulatory measurements are necessary in future work.

For both voiced and voiceless consonants, confusions were practically all within class. The intelligibility was somewhat higher for voiceless than for voiced phonemes.

For the vowels, we obtained a high proportion of correct identifications. This was an easy task, however, since in this case a forced choice paradigm was used in which the listeners were to choose among /i a u/. If the task had been to identify all Swedish vowels, the correct response rate would probably have been lower as suggested by ongoing analyses.

The listener panel mostly identified the normal speaker's consonants and vowels correctly. Interestingly, however, consonant identification was not perfect even in this very careful, but context free reading pronunciation.

There was a significant difference between the two listener groups when judging the consonants for the glossectomized speaker. The listeners who identified the consonants after the vowels performed better than the listeners who identified the consonants before the vowels ($p < .001$).

Table 1
Mean percentual intelligibility of consonants and analysis of errors on the basis of manner of articulation, place of articulation, and the voiced vs. voiceless feature.

	Mean intelligibility of consonants, %	Errors, %	
		Within class	Out of class
MANNER OF ARTICULATION			
plosives	42	52	6
fricatives	44	43	13
liquids	28	30	42
nasals	54	46	-
PLACE OF ARTICULATION			
bilabial	88	-	12
labiodental	85	-	15
dental	21	-	79
palatal/velar	9	-	91
laryngeal	100	-	-
VOICED VS. VOICELESS			
voiced	40	59	<1
voiceless	45	55	<1

Acknowledgments

The project is supported by The Swedish Council for Planning and Coordination of Research (Forskningsrådsnämnden) under grant 880252:3, A15-5/47 and The Swedish Cancer Society (Riksföreningen mot cancer - Cancerfonden) under grant 2653-B89-01X.

References

1. LINDBLOM, B. & SUNDBERG, J. Acoustical consequenses of lip, tongue, jaw and larynx movement. Journal of the Acoustical Society of America 50. 1971, 1166--1179.
2. MORRISH, L. Compensatory vowel articulation of the glossectomee: acoustic and videofluoroscopic evidence. British Journal of Disorders of Communication 19. 1984, 125--134.

Siket, nagyothalló és ép hallású tanulók beszédtempójának összehasonlító vizsgálata

Bujdosóné Arató Adrienne
Bárczi Gusztáv Gyógypedagógiai Tanárképző Főiskola
Fonetikai és Szurdopedagógiai Tanszék
Budapest, Bethlen Gábor tér 2, Magyarország

Beszédkutató laboratóriumunkban eszközfonetikai vizsgálatok segitségével siket, nagyothalló és ép hallásu tanulók beszédének üteme közötti különbséget vizsgáltam, a beszédtempó ugynevezett mikrostrukturáját elemeztem.

A vizsgálat középpontjában az a kérdés állt, hogy a hallási visszajelentésnek milyen szerepe van a beszédtempó alakulásában, illetve a hallássérülés milyen mértékben befolyásolja a beszédtempó alakulását.

Vizsgálataimat 5 ép hallásu tanuló / Hernád u. Ált. Isk. Budapest / 5 nagyothalló / Nagyothallók Ált. Isk. Budapest / 5 siket / Siketek Ált. Isk. Budapest / közremüködésével végeztem.
A tanulók életkora : 13 -14 év.
A hallássérült tanulók hallásveszteségének mértékét audiogramon szemléltetem jobb és bal fülre egyaránt, de az ép hallásu tanulókkal való összehasonlitás céljából az ő audiogramjukat is mellékelem.

Az anamnesztikus adatok alapján a siket tanulók hallássérülésének oka két esetben örökletes eredetü, két esetben csecsemőkori, ismeretlen eredetü virusfertőzés, és egy esetben a mater vesegyulladása miatti koraszülés /iker/ a hatodik hónapban.

A hallókészülék tipusa : PFONAK SUPER FRONT STEREO

A nagyothallók esetében egy szülési sérülést, egy ismeretlen eredetü és három örökletes eredetü halláskárosodást állapitottam meg.

A hallókészülék tipusa : WIDEX A 2T, WIDEX G2

Az elemzésre került egyszerü szöveg 4 mondatból és 35 szóból állt.

" Az egész család a rádió időjárásjelentését figyeli. Ilyemkor lehet eldönteni, milyen ruhába öltözzünk. Tanácsos-e esernyőt, esőkabátot vinni, vagy mehetünk-e strandolni ahogy terveztük? Télen érdemes-e szánkót, silécet magunkkal vinni ha a Bakonyba megyünk. "

Az előzetesen áttanulmányozott és felolvasott szöveg magnószalagra rögzitése után NOZOD tipusu oszcillográf segitségével oszcillogramok készültek. Ezek értékelését tempóindex számitások követték. A felvételek ugynevezett csendes szobában történtek.
Az oszcillogramot fényérzékeny papirra rögzitettük 5o mm/sec. sebességgel. Ez azt jelenti, hogy 1 mm 2o ms-nek felel meg. A beszédtempót tempóindex jelzi, meghatározott beszédszakasz folyamán. Kiszámitottam a szavak tempóindexét, melynek menete a következő.

A szó artikulálására forditott időtartamot elosztottam a szóban ejtett hangok számával. Ezután kiszámitottam, hogy az egy hangra esett időtartam -értékkel hány hangot lehet kiejteni 1 mp alatt. Igy közös alapra vonatkoztatjuk az artikulációs sebességet, és egy mutatószámot nyerünk - ez a tempóindex.

például: család ' ép hallás esetén /
hangok száma : 5
a szó hossza az oszcillogramon: 14 mm
a szó artikulálására forditott idő : 28o ms
az egy hangra eső idő: $\frac{28o}{5} = 36$

az egy mp alatt ejthető hangok száma : $\frac{1ooo}{36} = 17.85$

a " család " szó tempóindexe tehát :17.85

Mindhárom csoport szavainak tempóindexét kiszámitva, a nyert 525 adat alapján készültek el a tempóindex szóródási görbék, és kerültek elemzésre.

A siket tanulók 2.1 -16.oo-ig tartó tempóindex intervallumban ejtették ki a szavakat. Ebből egy szó sem fordult elő a 13.1 - 14.oo -ig tartó tempóindexen.
A szavak 85,14 %-a azaz 29.8∼3o szó került a 4.1 - lo.oo-ig tartó tempóindexekbe 3 fölötti előfordulási számban.

A nagyothallók 3.1 - 28.oo-ig tartó tempóindex intervallumban ejtették ki a szavakat, ebből egy szó sem fordult elő a 2o.1 -21.oo, a 23.1 . 24.oo, és a 26.1 - 27.oo-ig tartó tempóindexeken.
A szavak 83.1 %-a azaz 28.8∼29 szó került a 7.1 - 17.oo-ig tartó tempóindexbe 2 fölötti előfordulási számban.

Az ép hallásuak esetén: 8.1 - 31.oo-igtartó tempóindex intervallumban ejtették ki a szavakat, ebből egy szó sem fordult elő a lo.1 - 11.oo. a 25.1 - 26.oo, 26.1 -27.oo, 28.1 - 29.oo tartó tempóindexeken. A szavak 81.37%-a azaz 28.2 szó került a 11.1 - 22.oo-ig tartó tempóindexekbe 1 fölötti előfordulási számban.

Számszerüen a tempóindex átlagokat összehasonlitava az eredmények a következők:
siketek: 7.oo
nagyothallók: 11.42
ép hallásuak: 17.42

Eredményeink alapján a siketek szavainak tempója a hallókénak 4o.18 %-a, tehát megközelitően két és félszer lassabban olvasták fel a szöveget, mint a hallók. A nagyothallók szavainak tempója a hallókénak 65.55 %-a, tehát megközelitően másfélszer lassabban olvastak. A siketek szavainak tempója a nagyothallókénak 61.29 %-a.

Elemzéseim a hallássérültek beszédtanitásához kivánt metodikai utmutatást adni. Eredményeim alátámasztják a beszédtempóval való hangsulyosabb foglalkozás szükségességét az érthető beszéd kialakitása érdekében.

EXPERIENCES WITH THE G-O-H HEARING SCREENING METHOD

Zsolt FARKAS, Márta AMBRUS, Enikő NAGY, Jenő HIRSCHBERG, Erzsébet SIMON-NAGY

Outpatient Clinic, Érd, Hungary
and Heim Pál Hospital for Sick Children,
Dept. of ORL, Phoniatrics and Paedaudiology, Budapest, Hungary

INTRODUCTION

Hearing impairment is very common in the world. About 10% of the total population has a hearing loss. In the last decades, unfortunately little effort has been made all over the world, especially in the developing countries, to prevent this condition. It is estimated - according to the WHO General Assembly report, 1986 - that perhaps as much as 50% of the current hearing impairment could be totally avoided, or its consequences significantly reduced, by the application of appropriate measures (3). Such prevention should be introduced as an integral part of the health service in the world.

One part, and perhaps the most important of prevention, is screening. Our aim is to check the usefulness of the G-O-H system in the everyday clinical practice.

MATERIAL, METHOD, AND RESULTS

The G-O-H system uses synthesized words to assess hearing, speech perception and speech pronunciation of children. The usage of the set is very similar to a tape recorder. The inventors of the system generated - by means of a synthesizer and a computer - a special phonetically balanced monosyllabic word material for hearing screening purposes. The system is easy to use, no specialists are necessary for the measurements. The method can be applied between the third and fourteenth year, or more. Making the test you can obtain information on writing and reading

ability, as well. The portable set consists of the following: measuring device, headphone, special cassette with synthesized words, answer and picture sheets, writing pens, instruction and application manual. The evaluation is easy on the basis of the enclosed formula. The specially synthesized words contain only the necessary frequency components of each sound. The difference between natural and synthesized words is only in the redundancy of the frequency structure (1,2).

In another study we used this method to produce different types of hearing loss, to demonstrate hearing impairment.

The unskilled person, who makes the examination, has the key to the words, and can suspect the type and degree of hearing loss. Both the correct and incorrect answers are given on the sheets. The test contains altogether 44 Hungarian monosyllabic words. The average intensities are at 45 and 55 dB which can be chosen according to the background noise. The average intensity of the speech signal varies not more than 3 dB. Normal hearing listeners understand 90% or more, while hearing impaired persons only some of the words. The different categories of the answers are:

I. good hearing
II. slight hearing problem maybe, (10-20 dB),
III. hearing problem with great probability, (>20dB)
IV. hearing problem is certain; audiological examination is necessary as soon as possible (>30dB).

At the Outpatient Clinic of Érd (County Pest), social workers and paediatricians examined 130 children between the age of 3 and 6 years. At the same time, a parallel pure tone audiogram was determined with a screening audiometer. The tests were performed always in the same building and at the same time, before noon, to ensure stabile circumstances. Out of the 130 children 19, more than 10% had a hearing problem. The screening test was followed by ENT and audiological examination. In 6 children the only problem was earwax in the canal, in 12 serous otitis media was found. The G-O-H and pure tone audiometric results were checked by impedance audiometry, otoscopy and otomicroscopy. At the controll examination one child was lost, 2 was operated on because of enlarged adenoids and in one patient a middle ear ventilation tube was inserted in general anaesthesia.
4 children had dyslalia and dyslexia and were successfully treated by speech therapists.

We find the case report of a 5-year-old boy very interesting to be mentioned. On the right ear 90% of the words were repeated, while on the left only 10% of all of the words which were administered through the headphone. Only some of the sounds were correctly identified, the consonants were totally misperceived. At the clinical examination a conductive hearing loss was found both in the low and high frequencies, at 35 dB intensity. Appropriate therapy was performed and at the control examination hearing loss ceased, the small boy was cured.

DISCUSSION

The use of synthetic speech for screening purposes is a new and easy to apply method in the everyday practice. The G-O-H system provides a possibility to reveal qualitative and quantitative hearing disturbances, as well, because children have to percept sounds (coding - decoding), they have to store them and repeat the words. The process is much more sophisticated than to percept an 'unfamilier sound', i.e. pure tone and give a sign, as it is done at pure tone audiometry. In spite of these facts we think that the two methods are complementary and none of them sould be omitted. There are several advantages of the G-O-H method, last but not least the low costs. We hope that this new possibility in audiology gets his proper place in the near future.

REFERENCES

1. Gósy M., Olaszy G., Hirschberg J., Farkas Z.: Szintetizált szavak használata a beszédaudiometriában I. Elvi alapok és módszer (in Hungarian) Fül-orr-gégegyógy. 1985, 31, 92-96.
2. Gósy M., Olaszy G., Hirschberg J., Farkas Z.: Szintetizált szavak használata a beszédaudiometriában II. Klinikai alkalmazás (in Hungarian) Fül-orr-gégegyógy. 1985, 31, 227-233.
3. Prevention of Deafness and Hearing Impairment. 39th Gen.Assembly, WHO, 27 March 1986

SYNTHETIC SPEECH IN COMMUNICATION AIDS. EXPERIENCES IN SWEDEN

Karoly GALYAS, Elisabet ROSENGREN
Dept. of Speech Communication and Music Acoustics
Royal Institute of Technology, Stockholm, Sweden

Introduction

Speech synthesizers started to leave research laboratories during the seventies and found their first applications for the handicapped. The intelligibility of the speech wasn't very high but intensive research work has improved the speech quality. A synthesizer for Swedish was developed at the Department of Speech Communication, Royal Institute of Technology, Stockholm, Sweden. Being a text-to-speech system, it was capable of pronouncing any text.

Working in a small language community, the researchers directed their effort towards developing a synthesizer which not only could talk Swedish, but English and other European languages as well. Experimental systems were designed 1978 and one of the first systems was used in a special school for motorically handicapped children. The experiments provided useful information for the further development of the synthesis system which became commercially available 1982 from the Infovox company. An improved and smaller synthesizer was introduced in 1985. It could be equipped with eight different languages; and four voices were available. At that time the development of a portable and battery operated device for the handicapped started at the Royal Institute, using the Infovox system.

The speech is generated by rules in this system which is based on a small microcomputer and the program includes the most common exceptions of each language. Deviant pronunciation of certain words can be corrected and stored by the user and there are means for changing the stress of the sentence and the intonation.

In order to make synthetic speech available for Bliss symbol users a special device, called BlissTalk, was developed. Symbols can be selected either with a magnet or by scanning and special rules combine them automatically into grammatically correct sentences.

Applications

Experiments for application started at an early stage of the technical development: first with one person who was trained in special communication sessions (1978) and later, when synthesizers with small size became available, several others were involved. An evaluation study has shown that the quality of the synthetic speech was acceptable and several users benefited from a voice output communication aid (VOCA). However, very few systems were purchased for speech impaired because the synthesizer used was neither portable nor battery operated.

The development of a portable synthesizer was completed in 1986 in collaboration with the company Fonema AB and the device is marketed under the name Multi-Talk. It has all the features of the Infovox synthesizer and it can be equipped with four or five languages. It has a memory for abbreviations and storage for quickly accessible phrases and sentences. The Swedish Institute for the Handicapped bought five devices for the purpose of evaluation and a study was carried out during 1987.

Six speech-impaired persons between 7 and 58 years of age participated in the study. Most of them had cerebral palsy but two had other causes of their speech impairment. The test period lasted between one and two months.

The evaluation mainly focused on Multi-Talk as a communication aid, but educational applications were included, as two of the subjects had difficulties with reading and writing.

Each person was visited at least once and they had the continuous support of one person on site, in most cases a speech pathologist or a special teacher, who was responsible for the evaluation. This person was asked to fill in an evaluation protocol at the end of the test period.

The protocol was based on the IPCAS evaluation form. Some of the original items were excluded and others were added, e.g. questions dealing with intelligibility and attitudes towards the synthetic speech, preference for voice and the use of preprogrammed sentences.

The test period was a stimulating and positive experience for both the test subjects and others involved. Five of them had devices of their own prescribed during or shortly after the test period.

Multi-Talk is the first and only portable aid with synthetic speech that exists in Sweden. Most people have never listened to synthetic speech and consequently it was a new experience for both the speech impaired and others to discover the advantages of a speaking aid.

Some of the results that were reported in the protocol were that the user was able to:

- answer independently in class, also in a foreign language

- speak in a group

- use the telephone

- attract attention and initiate a conversation more easily

- have better control over the message through the auditory feed-back

- better express his/her personality

- prepare messages in advance

- speed up the communication by using preprogrammed sentences

The intelligibility of the speech was judged as good to acceptable, also on the telephone. The parents of one user, however, reported that they sometimes had difficulties in understanding the speech, especially at distance.

The attitudes and reactions towards the synthetic speech were mainly positive, in some case mixed. (Those who wanted to participate in the study naturally had a positive attitude from the beginning.)

Regarding the voice type, five of the subjects chose an age- and sex-adequate voice. One teenager girl who used the educational programs chose the male voice for its better quality.

Technical problems arose in three devices and the defective parts were exchanged.

A special program for educational purposes was loaded into the computer for the two pupils with reading and writing problems. This programme did not work in a satisfactory way, so these pupils felt that the device was unreliable. These faults were later corrected.

Despite these problems both the pupils, the speech pathologist and the teacher who used the educational programs found the training fun and stimulating.

An eight-year-old boy, who had almost unintelligible speech because of muscular disease, made great progress during the test period. The speech pathologist reported:

- increased interest in communicating

- improved power of concentration

- increased spontaneous speech and better sound production

He also learned to read during this period.

Based on this evaluation, the Multi-Talk was approved by the Swedish Institute for the Handicapped as a personal communication aid.

After the evaluation we have continued to follow these users, who have had their devices for a long time. We want to see how communication can function for skillful users, and what their future needs and wishes are.

It is also important that non-vocal persons who are not able to read and write can benefit from aids with synthetic speech. Some adaptations were made on the Multi-Talk for this purpose. Symbols and pictures, instead of letters, were attached to the keys, and words and sentences were programmed into the "quick store memory." A few mentally retarded adults and a non-speaking preschool child have used this adapted aid, and for the first time they have had the possibility to express a word or a sentence vocally.

For Blissymbol users, a special device called BlissTalk has been developed. The user has access to 500 symbols, either by direct selection with a magnet or by row-column scanning. The symbol board is programmable, and the 500 symbols can be selected from a vocabulary of 1500 words. Special grammar rules make automatic corrections resulting in grammatically correct sentences. Available languages are Swedish, English, French and Spanish. The BlissTalk is approved as a communication aid.

SPEECH PERCEPTION PERFORMANCE: EVIDENCE FOR OR AGAINST A HEARING AID

Mária GÓSY
Institute of Linguistics, Hungarian Academy of Sciences
Budapest, Hungary

In general, it is relatively unproblematic for the audiologist to decide whether a child having hearing difficulties should be aided or not. Audiological examinations, in particular pure-tone and speech audiometry, provide sufficient data for a decision concerning the type and acoustic quality of the hearing instrument as well (1). Pure-tone audiometry is the traditional way of assessing hearing ability, whereas speech understanding level is normally the main criterion for defining the required acoustic properties of a hearing aid. However, there are many cases when, for various reasons, the pedoaudiologist cannot be convinced whether a hearing aid will be of great help or, on the contrary, will disturb the child's perceptual mechanism. Besides, in view of the social disadvantages that a hearing aided child is likely to encounter in everyday life, and the price of an instrument, the solution to the problem -- whether the child ought to be aided or not in marginal cases -- should not be based on a subjective decision of the pedoaudiologist. Scientific knowledge and clinical experience do not enable the audiologist to decide correctly in case the child's co-operation seems to be vague and/or his performance in pure-tone and speech audiometry is in complete contradiction.

In order to predict the subjective acoustic quality of an instrument with some degree of accuracy, its objective performance must first be clearly specified. However, in many cases the speech understanding of the aided child does not improve as much as it has been expected or the understanding process is reported to be disturbed by the amplified noises.

These experiences led us to the conclusion that the usual audiological examinations (including speech audiometry) do not provide sufficient evidence in every case for or against a hearing aid. In solving this problem, the actual speech perception and speech understanding level of the hearing-impaired child should be assigned crucial importance. Accordingly, an age-specific test-package for assessing the speech perception level of normal-hearing children has been developed in order to define age-specific criteria for 'correct' speech perception of the mother tongue. Efforts have also been made to obtain information on the operations of each hypothetical level quasi-separately, ie. to detect which (if any) of the decisions the understanding mechanism has to perfom are mistaken or incorrect.

Method and material

The test-package consists of 10 subtests; their speech material varies from isolated words through sentences up to a longer text. These speech materials have been manipulated by various methods (such as masking by white noise, speeding up, and frequency filtration). Natural Hungarian speech announced by a trained male speaker and also artificially generated synthesized speech have been used for the subtests. Some of the listening tests have been administered to the subjects through headphones, others through a loudspeaker in a silent room. The intensity level of the recorded materials was controlled by the G-O-H device and by an audiometer.

350 normal-hearing and 50 hearing-impaired children (ages between 4 and 10) have been examined with the test-package. The hearing-impaired children seemed to be 'marginal cases' after their clinical measurements. Some of them showed a very poor performance in clinical speech audiometry which would not have been expected after their pure-tone audiometric results. The other part of these children had a quite severe hearing loss (according to the pure-tone audiometry), however, they rejected to accept a hearing aid because they had not experienced failures in their everyday verbal communication.

Results with the test-package

On the basis of the results with the normal-hearing children, a special diagram of speech perception development has been set up, in terms of which the perception performance of the hearing-impaired children could be "objectively" compared with that of normal-hearing ones.

1. The first examination was performed by means of the G-O-H hearing screening device. Identification of synthesized monosyllables -- separately administered to the right and left ear, at the intensity levels of 45 and 55 dB -- revealed the deviation of hearing-impaired children from the average performance of the normal-hearing population of the same age. Degrees of deviance from normal responses gave us information also about the perception level of the examined child. On the basis of the responses, the hearing threshold curve could be drawn, which was important particularly in cases when the child's performance in pure-tone audiometry was vague.
2. In everyday communication the spoken message is frequently covered by noises of various types and intensities. For successful communication to take place the speech understanding process should work correctly even under noisy circumstances. The 'coctail-party problem' arises as a real problem particularly for children because they do not have as much practice in understanding speech as adults do. The second task of the test-package was to identify/understand 10 well-formed sentences masked by white noise. The signal/noise ratio is 4 dB. The average intensity levels (through a loudspeaker) were 60 and 70 dB.
3. Word-identification was examined by 10 (mono-, bi-, and trisyllabic) words also masked by white noise. Slight perceptual differences among sentences and words provided us with reliable information on the operations of the acoustic and phonetic levels of the understanding mechanism.
4. Identification of fast sentences (the normal tempo was speeded up by means of a Varyspeech by 30 % of the original version) gave us an opportunity to detect basic hearing and perceptual problems in decoding a speech signal. The first signs of a disturbed speech perception/understanding process very often appear when the process is forced to work in a narrower time structure.
5. The next subtest was the identification of sentences filtered by pass-band filtration with the slope of 36 dB. After filtration all sentences were sounded in the frequency range of 2200 to 2700 Hz. There were normal sentences without any filtration announced originally by the same male speaker for control. In many cases of hearing impairments, speech was identified more correctly in a narrow frequency range than in its whole acoustic structure.
6. The next subtest was performed like a "face-to-face" game between the child and the examiner. 10 animals' names had to be guessed by the child from watching the exact (but silent) lip-articulation of the word in question by the examiner. The results show an inverse correlation between lip-reading ability and perception. The better the child's lip-reading

ability, the more likely that he has a hearing loss.

7. The acoustic/phonetic levels were further examined by the next task involving five nonsense sound-sequences of two, three, and four syllables which partly met, partly contradicted the Hungarian phonotactic rules. They were pronounced by the examiner and had to be repeated by the child in a correct way.

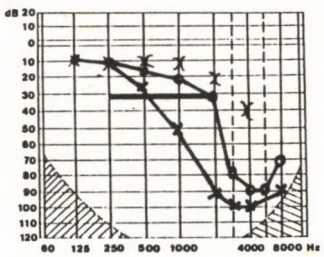
8. The so-called 'central hearing deficiences' were detected by a special dichotic test. This test material contained 5 bisyllabic non-compounds and another set of 5 four-syllable compound words. These clearly recorded words were electrically separated at their mid-point (like Mon-day), then they were recorded so that the first part of the word could be administered to one ear, while the other part of the same word at the same time to the other ear. The child's task was to "synthesize" the two separated parts of each word. If she/he was successful in understanding the whole word, it could be supposed that his/her central hearing mechanism worked well.

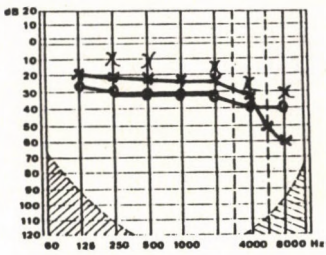
9. Verbal and visual short-term-memory examination was performed by displaying 12 words and 12 color pictures. The child had to recall items that he/she had heard/seen. In interpreting the results, not only the number of recalled items but also the order of recalling was taken into consideration.

10. Finally, a short story was told to the child by the examiner. During the story telling they simulated a conversation situation, so that the child could use some para- and extralinguistic features for comprehension. The comprehension of the story was checked by questions to be answered by the child. Responses to the carefully prepared questions highlighted (i) the level of associations, (ii) the differences between the perceptual mechanisms of children and adults, and (iii) the 'speech reading' ability of the child. Comparing the results of a hearing-impaired child with those of his age-group, an acceptable decision could be made for or against the hearing aid. Let us give two brief examples (audiograms and performance).

V.R.
(girl)

age 14





S.P.
(boy)

age 12

Speech perception performance of two examined children

Tests	V.R. (girl)	S.P. (boy)
correct identification of sound-sequences	50 %	25 %
correct working of the acoustic/phonetic/ phonological levels	80 %	40 %
identification of filtered/normal sent.	0 %/100 %	40 %/40 %
identification of fast sentences	100 %	70 %
dichotic test	90 %	30 %
lip-reading	100 %	30 %
verbal/visual memory	good	good
comprehension of text	100 %	20 %

On the basis of these results, a decision was made against the hearing aid in the case of the girl while a hearing aid was recommended to the boy. These children (together with the others from the total of 50) were subsequently re-examined after 1 year. The correct working of the test-package (Fig. 1) has been supported: no one has been found for whom the former decision had to be changed.

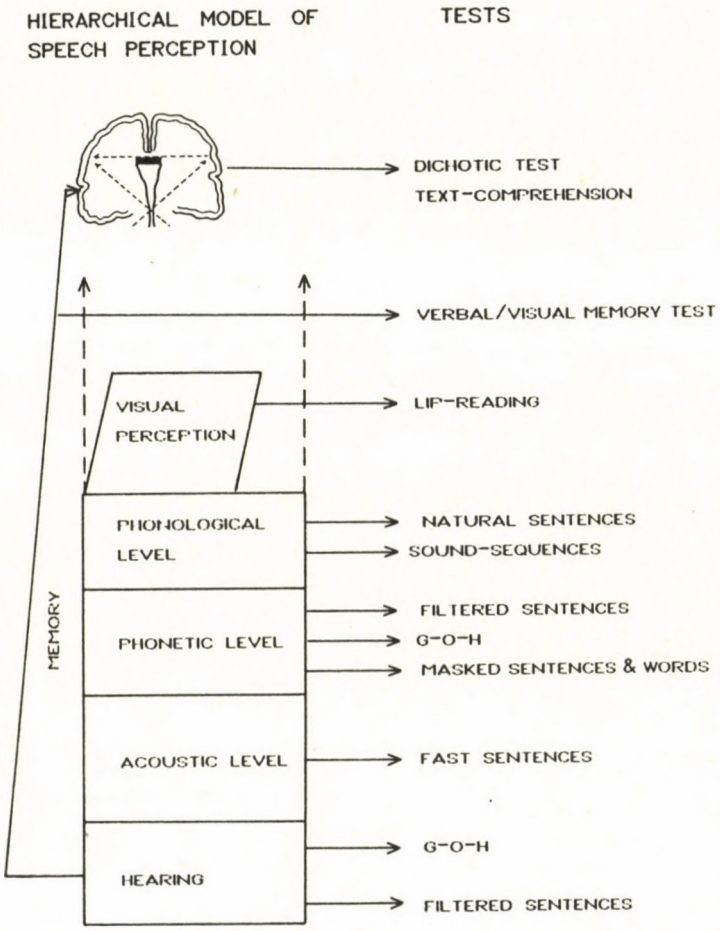


Figure 1.
The structure of the test-package

Reference

1. BERGER, K.: Questions about hearing aid prescription. Kent 1981.

DIE BEURTEILUNG DES VERHÄLTNISSES ZWISCHEN SING- UND SPRECHSTIMME ANHAND EINER QUANTITATIVEN MESSMETHODE

Tamás HACKI
Klinik und Poliklinik für Phoniatrie und Pädaudiologie
der Medizinischen Hochschule Hannover / BRD
(Prof.Dr.E.Loebell)
MAV Központi Kórház és Rendelöintézet, Budapest

Eine Diskrepanz zwischen dem Sprechstimm- und dem Singstimmbefund gehört in der phoniatrisch-logopädischen Praxis nicht zur Seltenheit. Annehmbare Singstimmleistungen bei schlechten Sprechstimmleistungen sind gelegentlich bei Stotterern, Aphasikern, Parkinson-Kranken, bei Patienten mit spastischer Dysphonie und nicht zuletzt bei Patienten mit funktionellen Stimmstörungen, zu beobachten.
Auf dem mehr oder weniger intakten Vorhandensein der Singfähigkeit bei gestörter Sprache bzw. Sprechstimme basieren verschiedene Therapiemethoden.
Unsere Absicht mit dieser Arbeit ist es, sich mit dem Verhältnis zwischen Singstimme und Sprechstimme, sowie ihren Eigenschaften anhand einer quantitativen Messmethode auseinanderzusetzen im Intresse der Diagnosestellung als auch der Bestimmung der Therapieziele.

Die zweidimensionale Darstellung der quantitativen Leistungen der Singstimme, das Phonetogramm" von C.H. Waar und P.H. Damste (1968), sowie P.H.Damste (1970) befindet sich in kontinuierlicher Weiterentwicklung: H.K. Schutte (1980), M. Gross (1980), F. Klingholz und F.Martin (1983), Komiyama und Mitarb. (1984), Pedersen und Linskow Hansen (1986), W.Seidner, J. Wendler, H. Wagner, A.Rauhut haben für eine bessere diagnostische Aussage und für die technische Entwicklung der Stimmfeldmessung gesorgt.
Seidner (1985) stellt fest, daß die Beurteilung der Sprechstimme anhand des Singstimmfeldes trotz fester Beziehung zueinander nicht optimal ist. Um zu einer besseren Beurteilung zu gelangen, erstellt er "Sprechstimmprofile". Die Probanden werden zum Zählen aufgefordert. Die Sprechfrequenzen werden nach dem Gehör des Untersuchers an der Klaviatur des Stimmfeldmessgerätes ausgewählt, und der Schallpegel gemessen. Die so gewonnenen Schallpegelmaxima werden in das Stimmfeld als Sprechstimmprofile eingezeichnet.
Graming(1988)untersucht die Sprechstimme während des Lesens eines Textes mit verschiedener Lautstärke. Es werden sowohl die Grundfrequenz als auch ein zeitlicher Durchschnitt des Schalldruckpegels (Leq) der Stimme nachträglich analysiert und als Trajektorien in das Stimmfeld eingezeichnet.
Aufgrund der Ergebnisse der Untersuchungen von 18 Probanden und 20 Patienten hält die Autorin es für sinnvoll, die Sprechstimme der stimmgestörten Patienten im Verhältnis zu ihren Stimmfeldern zu analysieren: sie hält es für angemessener, über optimale Trajektorien im Stimmfeld als über eine optimale Stimmhöhe zu sprechen.

Zur Analyse der Sprechstimme erscheint uns eine zweidimensionale Darstellung der quantitativen Sprechstimmleistung empfehlenswert zu sein: Als Analogie zum Stimmfeld (= Singstimmfeld) bietet sich die Verwendung des Begriffes "Sprechstimmfeld" an.

Material und Methodik

Das Prinzip des Stimmfeldmeßcomputers "Phonomat" haben Angster, Hacki und Taba in den Jahren 1984-85 erarbeitet und in den folgenden Jahren mit der Fa. Homoth (Hamburg) weiterentwickelt. Im Gegensatz zu herkömmlichen Stimmfeldmeßgeräten kann der Patient auch ohne Tonvorgabe phonieren, wobei der Phonomat automatisch Tonhöhe (Grundfrequenz, F0) und Lautstärke (SPL, dB lin.) mißt und diese auf dem Bildschirm zweidimensional (F0 X-Achse in Halbtonschritte, SPL Y-Achse in dB) darstellt. Die Meßwerte, in Form von Punkten, erscheinen direkt am Bildschirm des Stimmfeldcomputers. Die visuelle Kontrolle erleichtert die Untersuchung von musikalisch ungeübten Patienten erheblich. Auch beim Untersucher ist eine Musikalität keine Voraussetzung. Bei der Untersuchung von stimmgeübten Personen können Töne auch über Lautsprecher vorgegeben werden. Nach der Untersuchung wird das Stimmfeld auf Befehl umrandet, der Flächeninhalt berechnet und in den Diskettenspeicher eingelesen.
Bei der Aufzeichnung der Singstimmproduktion erzielen wir die Erfassung der physiologischen Singstimmgrenzen.
Nach der Aufzeichnung des Singstimmfeldes bestimmen wir den vollen Umfang und die gesamte Dynamik der Sprechstimme als Sprechstimmfeld: Zur möglichst umfassenden Sprechstimmproduktion wird der Patient mit Hilfe von Vorstellungsbildern motiviert.
Es werden Zahlen, sinnlose Silben (möglichst nasale-vokale-Verbindungen), einfache, kurze Sätze leise, mit Umgangslautstärke und schließlich laut phoniert, im Anschluß daran wird gerufen.
Singstimm- und Sprechstimmfeld können anschließend übereinander auf dem Bildschirm dargestellt und mit deren numerischen Vergleichswerten tabellarisch ausgedruckt werden.
Untersucht wurden 105 Stimmgesunde und 97 Stimmkranke mit Stimmstörungen ohne organichen Befund.

Befundergebnisse

Die Untersuchung von 105 stimmgesunden Personen ergibt folgendes:

a) Im Normalfall liegt das Sprechstimmfeld im unteren Drittel des Singstimmfeldes. Abb.1.

b) Die Dynamikgrenzen des Sprechstimmfeldes nähern sich den Dynamikgrenzen des Singstimmfeldes oder erreichen sie.Im entsprechenden Frequenzbereich kann gleicherweise leise und laut gesprochen wie gesungen werden. Abb.1.

c) Die Rufstimme, als eigenständiges Rufstimmfeld, liegt im Frequenzbereich des Registerüberganges . Sie kann in der Dynamik über das Singstimmfeld hinausreichen. Abb.1.

Die Untersuchungen von 97 Stimmkranken ohne organnischen Befund (funktionelle Stimmstörungen) ergaben folgende Ergebnisse:

Das optimale Verhältnis zwischen Sing- und Sprechstimmfeld kann in solchen Fällen in verschiedener Weise gestört sein.

1.1 Das Rufstimmfeld fehlt. Der Patient ist nicht in der Lage, eine Rufstimme zu bilden (in 15 Fällen).
1.2 Das Rufstimmfeld nähert sich nicht der oberen Dynamikgrenze des Singstimmfeldes. Es wird eine schwächere Ruf- als Singstimme gebildet (in 6 Fällen).
1.3 Das Rufstimmfeld befindet sich im Bereich des Kopfregisters. Die Stimme des Patienten kippt beim Rufen ims Kopfregister (in 4 Fällen).
1.4 Das Rufstimmfeld plaziert sich in der Einbuchtung (Registerübergang bzw.Registerbruch) des Singstimmfeldes (in 7 Fällen).
1.5 Das Rufstimmfeld liegt oberhalb des Singstimmfeldes. Die Singstimme ist wesentlich schwächer als die Rufstimme (in 13 Fällen).

Bei guten oder annehmbaren Singstimmleistungen kann sich das Sprechstimmfeld -als Ausdruck einer gestörten Sprechstimmleistung - eingeschränkt und/oder falsch plaziert darstellen.

2.1 Das Sprechstimmfeld ist in Richtung oberer Dynamikgrenze verlagert. Der Pat. kann leise singen, spricht aber gewohnheitsmäßig laut (in 3 Fällen).
2.2 Die obere Dynamikgrenze des Sprechstimmfeldes ist nach unten verlagert. Der Pat. hat eine zurückgenommene, nicht genügend steigerungsfähige Sprechstimme (in 16 Fällen).
2.3 Das Sprechstimmfeld ist in Richtung höherer Frequenzen verlagert. Die mittlere Sprechstimmlage ist zu hoch und/oder die Sprechstimme wird mit zunehmender Lautstärke zu hoch (in 7 Fällen).
2.4 Das Sprechstimmfeld ist in Richtung tiefer Frequenzen verlagert, reicht gelegentlich über die untere Frequenzgrenze des Singstimmfeldes hinaus. Die Sprechstimme ist "nach unten gedrückt", zu tief (in 6 Fällen).

Bei erhaltenen quantitativen Sprechstimmleistungen können die quantitativen Singstimmleistungen in verschiedener Weise eingeschränkt sein.

3.1 Das Kopfregister der Singstimme fehlt oder es ist stark eingeschränkt (in 9 Fällen).
3.2 Die obere Dynamikgrenze des Singstimmfeldes ist nach unten verlagert; der Patient kann nur leiser singen als sprechen (in 2 Fällen).
3.3 Die untere Dynamikgrenze des Singstimmfeldes liegt höher als die des Sprechstimmfeldes. Es kann leiser gesungen als gesprochen werden (in 5 Fällen).
3.4 Das Singstimmfeld ist in jeder Richtung eingeschränkt (in 9 Fällen).
4. Singstimm- und Sprechstimmfeld sind in ähnlicher Weise eingeschränkt (in 17 Fällen).
5. Singstimm- und Sprechstimmfeld sind dissoziiert. Diesen Fall konnten wir bei der Mutationsstimmstörung beobachten (in 4 Fällen).

Diskussion

Die Plazierung des Sprechstimmfeldes im unteren Frequenzdrittel des Singstimmfeldes bestätigt die bisherigen Erfahrungen aus der wissenschaftlichen Forschung und der Praxis.
Die Ausdehnung des Sprechstimmfeldes in Richtung der Dynamikgrenzen des Singstimmfeldes, also die Ausnutzung der maximalen physiologischen Gegebenheiten für die Sprechstimmproduktion, scheint von der psychomotorischen Steuerung der Sing- bzw. der Sprechstimme abhängig zu sein. Im Optimalfall kann der Proband seine ganze Stimmdynamik auch für die Sprechstimmproduktion einsetzen.
Im Gegensatz dazu haben wir bei einigen - im klinischen Sinne stimmgesunden, aber aus stimmpädagogischer Sicht ungünstig veranlagten Probanden - hauptsächlich aber bei stimmgestörten Patienten, wechselnde Beziehungen zwischen Singstimm- und Sprechstimmleistungen beobachten können und haben Diskrepanzen festgestellt.
Eine schwache Singstimme bei normaler Sprechstimmdynamik (siehe 3.2) scheint eher eine Frage der fehlenden Singstimm-Erfahrung zu sein, als eine Diskrepanz in der psychomotorischen Steuerung.
Das Phänomen, leiser sprechen als singen können (siehe 3.3), wird auch bei professionellen Sängern beobachtet, die auch Piano-Töne gestützt phonieren.
Die Lage des Rufstimmfeldes steht in enger Beziehung zum Übergang vom Brust- zum Kopfregister. Die Rufstimme repräsentiert die lautesten und die höchsten Anteile der Bruststimme. Die Plazierung des Rufstimmfeldes im Singstimmfeld kann daher der Bestimmung des "natürlichen Brustregisters" dienen.
Gelegentlich reichen das Sprechstimmfeld und das Rufstimmfeld über das Singstimmfeld hinaus. Daher soll man neben dem Singstimmfeld auch ein Sprechstimmfeld darstellen, um die Dynamik-und Umfangsmaxima der Stimme unter ihrer gegenwärtigen physiologischen Bedingungen zu erfassen.
Es scheint, daß Diskrepanzen zwischen Sprech- und Rufstimmfeld einerseits und Singstimmfeld andererseits, auch auf stimmliche Reserven hinweisen können (siehe auch unter 1.4 , 1.5 und 3.2).
Zusammenfassend können wir feststellen, daß die aufgezeigte Methode geeignet ist, physiologische Daten der Sprech- und Singstimmbildung objektiv zu erfassen und miteinander zu vergleichen.

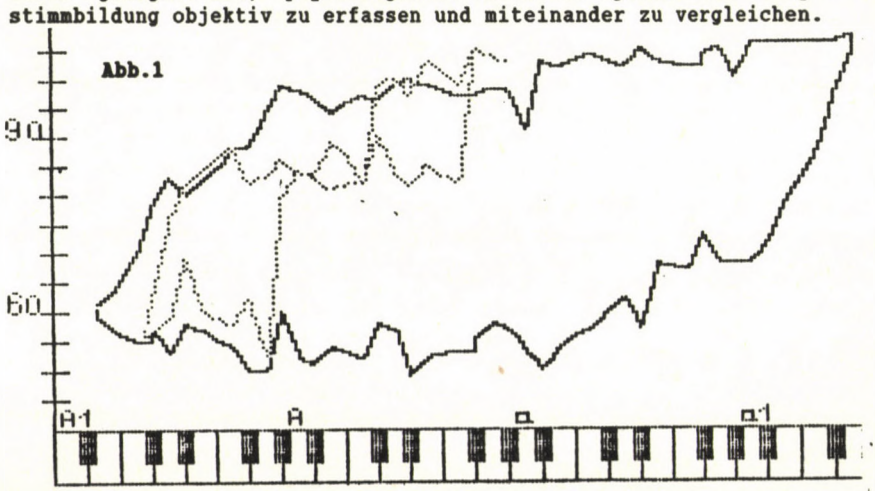


Abb.1

THE PERCEPTION AND PRODUCTION OF A VOICING CONTRAST IN PROFOUNDLY HEARING-IMPAIRED CHILDREN

Valérie HAZAN, Adrian J. FOURCIN, Evelyn ABBERTON
Dept of Phonetics and Linguistics,
University College London

Introduction

The ability to perceive the voicing contrast in plosives auditorily is particularly important for hearing-impaired children as they are not marked visually. This ability has been assessed in a number of studies of moderately to profoundly hearing-impaired children (eg Parady et al., 1981, Johnson et al, 1984, Abberton et al, 1987). These studies have found that while moderately hearing-impaired children are able to label stimuli along a voicing continuum in an essentially normal way, severely-to-profoundly hearing-impaired children vary widely in their ability to label the contrast. The majority of studies present average results obtained from a "homogeneous" group of children. Given the sizeable differences in rate and level of development attained by hearing-impaired children of similar age and educational background (Hazan and Fourcin, 1985), averages mask quite variable individual results. In the study presented here, an unselected group of seventeen children, all born within a 12 month period and educated in a hearing-speaking environment, were tested over a three year period. Both their perception and production of a small number of speech contrasts were assessed approximately three times a year. Results are presented here for six of these children matched in three pairs in terms of their pure tone audiogram average hearing loss (3 FA at .5 kHz, 1 kHz, 2 kHz in better ear) , age and educational background.

Subjects

The six children, aged between 7 and 8 at the beginning of the testing period, were all pupils at the Birkdale School for Hearing-Impaired children. The two least impaired had average losses of 83 dB HL and 90 dB HL (Child 1 and Child 2); the two most impaired had average losses of 112 dB HL and 115 dB HL (Child 16 and Child 17), and two had average losses of 102 dB HL (Child 7 and Child 8). Of this group, four children were congenitally hearing-impaired; Child 7 was deafened at 3;6 years and Child 8 was deafened at 0;4 years.

Test procedure

The minimal pair COAT-GOAT was used in the investigation of the perception and production of the voicing contrast in initial plosives. Interactive synthetic speech identification tests were used for the perceptual assessment. Children were tested on their identification of stimuli along a voicing continuum varying in Voice Onset Time (from 20 to 100 ms) and F1 onset frequency. They were tested on a maximum of ten occasions between March 1985 and June 1988.Stimuli and test procedure have been described in more detail in Hazan and Fourcin (1985). Results are presented in the form of individual identification functions (see Figure 1). Identification functions for 1985 were averaged from results obtained at the March-November 1985 sessions. The 1988 results were obtained at the June session.

Production measurements were based on simultaneous microphone and laryngographic digital recordings made at each testing session . The laryngograph allows non-invasive monitoring of vocal fold activity during normal speech production. Children were recorded while naming pictures representing words forming minimal pairs and while producing spontaneous speech. The waveforms obtained were digitised on a mini-computer and measurements were made of the Voice Onset Time (taken from burst release to the first regular vocal fold vibration).

Perception of voicing contrast

Identification functions obtained for categorically perceived contrasts are typically S-shaped and characterised by areas of consistent labelling at the endpoints of the range, and a sharp gradient around the phoneme boundary. During speech development, before the categorical perception of a given contrast is achieved, children go through a stage of "random" labelling, with inconsistent responses to endpoint stimuli, and then of "progressive" labelling, with consistent labelling of endpoint stimuli but inconsistent or continuous labelling of intermediary stimuli (Simon & Fourcin, 1978). In 1985, categorical functions were obtained for Child 1 only. The other subjects were quite inconsistent in their labelling and were giving less than 100% correct responses to the endpoints of the range. Child 17 was not tested on this contrast because he was totally unable to label the endpoint stimuli in an initial training stage.

Three years later, progress is seen for most children. Children 1 and 2, who are the least impaired are labelling the COAT-GOAT contrast in a categorical manner. Child 7 is also labelling the contrast categorically while Child 8, with the same pure tone average loss but congenitally hearing-impaired, is still labelling the endpoints of the range at random. Child 16 is labelling the voiced stimuli consistently but the voiceless stimuli randomly. The greatest development is seen for Child 17, who even with a pure tone average loss of 115 dB HL, was able, by the age of ten, to label the endpoints of the contrast consistently. He was still, however, labelling the intermediate stimuli at random.

Production of voicing contrast

1. General comments

The plosives /k/ and /g/ were produced in the "goat-coat" minimal pair but were also contained in other words of the Edinburgh Articulation Test which was administered at each session. A first assessment of whether these segments are present and contrastive in 1985 and 1988 is therefore made from phonetic transcriptions of the productions of the words "coat" and "goat". A summary is presented in Table 1.

Table 1:

Child	1985		1988		
	/g/	/k/	/g/	/k/	
1	[g]	[k]	[g]	[k^h]	
2	[g]	[k]	[k^h]/[$\underset{\smile}{d}$]	[k^h] / [$\overset{\smile}{g}$]	variable
7	[g]	[k]	[g]	[k]	
8			[$\underset{\smile}{d}$]	[k^h]	
16			[gd]	[ɦ]?*	
17			[d]	[$\underset{\smile}{d}$]/[ʔ]	

* Note: in 1988, Child 16 produced a sound perceptually intermediate between [d] and [g]; not like a palatal however. It is possible that it is a double articulation [gd].

Some children (child 2,8,16,17) could not produce one or both segments in the test words but did produce them in other words of the Edinburgh Articulation Test , sometimes in intervocalic or word-final position.

2. VOT measurements

Recordings made at the November 1985 and June 1988 sessions were analysed, and the Voice Onset Time, measured from the start of the plosive burst to the onset of closure for the first regular vocal fold vibration in the laryngographic waveform, calculated for the first production of the COAT-GOAT minimal pair.In some cases (see above) children produced tokens with appropriate voicing but inappropriate place of articulation. The results are presented below.

Table 2: Voice Onset Time values in milliseconds for tokens of the initial plosives in "goat" and "coat"

VOT in ms

	[g]	[k]	[g]	[k]
Child	1985		1988	
1	9	99	12	75
2	8	64	11	100
7	14	75	11	71
8	-		27	60
16	-		8	-
17	-		-	
Normally-hearing children				
Average VOT measures (Simon, 1976)				
7 years	15	68	10 years 34	104

Discussion

The aim of this ongoing longitudinal study was firstly to look at evidence of natural development in the ability to perceive and produce a voicing contrast in initial plosives over a three year period, and secondly to look at the relation between the ability to perceive and produce this contrast in individual hearing-impaired children.

In 1988, a certain correspondence can be found between a hearing-impaired child's ability to perceive and produce a voicing contrast in initial velar plosives. This ability is not directly related to the pure tone audiogram average as can be seen by the results. Children 1 and 7 can both perceive and produce the voicing contrast in an essentially normal manner. Child 2 (PTA: 90 dB HL) is labelling the contrast in a categorical manner but still producing the contrast inconsistently, with the plosive /g/ not yet fully acquired. Child 8 is neither perceiving nor producing the contrast consistently. Child 16 is labelling voiced tokens consistently, but not voiceless tokens. He is producing a voiced sound perceptually intermediate between [d] and [g], and [h] for the voiceless token. Child 17 is labelling the stimuli progressively, and is therefore still in the process of acquiring the contrast. In terms of production, again, some contrast is made between voiced and voiceless tokens, but the contrast is still immature.

In all children, progress has been seen over the three year period. These developments appears to be the result of natural maturation, and not of specific training, as none was given, or of habituation to the test, as the children developed at different rates and stages of the testing period. At the age of ten, whilst development of perception and production of contrasts is virtually complete in normally-hearing children, development is still being seen in the acquisition of a basic speech contrast in hearing-impaired children.

References

1. ABBERTON,E, HAZAN,V. & FOURCIN,A.J. (1987) Speech pattern acquisition in profoundly hearing-impaired children. Proceedings of 11th International Congress of Phonetic Sciences, Tallinn, USSR

2. HAZAN,V, and FOURCIN, A.J. (1985) Microprocessor-controlled speech pattern audiometry:preliminary results. Audiology,24,325-335.

3. JOHNSON,D., WHALEY,P. & DORMAN,M.F. (1984) Processing of cues for stop consonant voicing by hearing-impaired listeners. J. of Speech and Hearing Res., 27, 112-118

4. PARADY,S., DORMAN,M., WHALEY,P. & RAPHAEL,E. (1981) Identification and discrimination of a synthesised voicing contrast by normal and sensorineural hearing-impaired children. Journal of Acoustical Society of America, 69, 783-789.

5. SIMON,C (1976) A developmental study of acoustic pattern production and perception in voiced-voiceless oppositions. PhD thesis, University College London

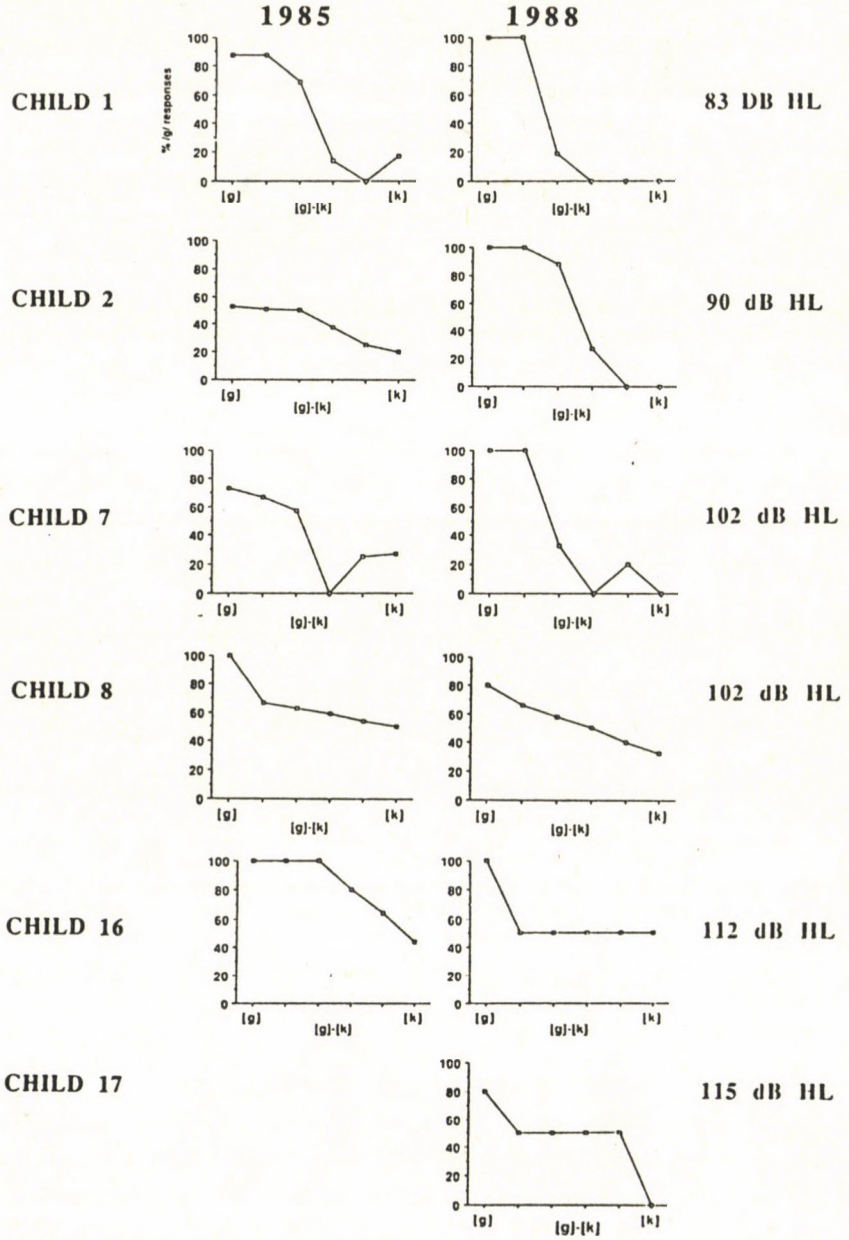


Figure 1: Identification functions for GOAT-COAT contrast

FROM SIGN TO TEXT: HIERARCHIC LINGUISTIC RESTRUCTURING IN GLOBAL APHASIA

A. HEGYI, speech therapist
Z. JANKA, neuropsychiatrist

Szent-Györgyi University Medical School,
Department of Neurology and Psychiatry,
Szeged, Hungary

Since the 1970-es a new pedagogic method aiming to develop the language ability of normal children of nursery age and that of children with developmental and other language disorders has become available /3,6,13,14,15,16/. We had the opportunity to implement it into practice with the consideration of improving the details and also rehabilitating linguistically the adult patients suffering from global aphasia. It is important to note that all the strategic units of the method can be demonstrated in patients with global aphasia.

Before formulating our therapeutic concepts we have to outline the ideas about language disorders related to brain dysfunction. The ideas are based on our own experiences as well as on literature date of the recent years /1,2,18/. Two types of disorders related to brain damage can be distinguished: /i/ the disorder of the symbol system - aphasia, and /ii/ the disorder of the non-symbolic system - agnosia and apraxia. Aphasia results from dysfunctioning of a central integrative mechanism leading to the loss of the patient's ability to develop associations generated by new stimuli and based on previously acquired knowledge. Normally, these associations promote recalling the symbol. In this way, it is not the language symbols stored earlier are damaged but the power to recall them.

Our concept is that the lack of language ability in patients with global aphasia should be seen as a symbol use disorder. Consequently, the goal in the therapeutic process is to enable the patients to differentiate the simplest conventional signs, to construct their most suitable sequences so as to use this more and more complex arragement with success as means of linguistic communication at the different levels of the language system.

At the beginning of the therapy /first phase/ the patients can neither understand nor speak. The dysfunction of the recalling processes suggests that they suffer from a general language disorder. They do not react to the stimulus of direct therapy as the disorder of symbols obstructs both comprehension and expression processes. The introduction of complex indirect stimuli must be considered as most effective /7,17/. Disorders of symbols in writing and reading appear in a similar way as those in speaking and comprehension /11/. We help the patients acquire the ability of writing and reading simultanously which ensures that the perception of language symbols takes place in a modulative, mutually stengthening way.

It is important to emphasize that if the patient's state does not allow a proper psychometric assessment of the symptoms, a detailed and exact description about the capacities of the person is required which includes anatomical, speech-specific, psychic, and psychosocial factors as well.

The preparatory phase can be called as activating phase in which we strengthen the recognising and spontaneously designating function of the language by joint acting, acting directed by strong gestures and by verbal supports connected to the former ones. First we make the patient imitate analogue-acting sentences to promote concept-thinking. During this time the subject recognises a form similar to the acting /e.g. a circle/ he or she could choose it out of other forms, and to classify the similar ones. Later the subject is able to classify the forms not only by their size, colour, and surface, but by considering several points of view at the same time. Comprehension is extended on the spatial position of the forms, too. The same procedure is applied to analyse and synthetise various objects and object pictures of the environment. This corresponds the development of preverbal learning ability /9/.

The activating phase is followed by syndrom-specific exercises supporting the main goal of our therapy concept. In other words, we enable the patients to differentiate the meaningful elements of language system on different levels, and to structure their correct sequence in order to communicate either in the form of speaking, reading, or writing. The experiments are linked to the following levels: /i/ sound formation, /ii/ word-structure, /iii/ sentence structure, and /iv/ text-structure.

Spiritual efforts are needed to make the articulated sounds to be adequate for the presentation of thoughts. Graduation is a basic concept in this method. If the subject is able to classify forms or objects into definite groups on the ground of their different properties, then his or her registration memory is able to input and store several, simultanous stimuli. This ability is used to help the patient recall language symbols connected to sound-forming. We assume that the recall of vowels as musical elements is helped by the subdominant cerebral hemisphere as well /8,10,12/. So, on the level of sound- formation the sounds of á [ɑ:] , i [i, i:] , u [u, u:] found in all the languages, as well as the frequent a [ʌ] vowel in the Hungarian language are to be differentiated and picked out of the sound sequences by the patient when joint visual, kinetic, and tactile stimuli are applied. The sound is pronounced, its letter symbol is shown, and we follow our joint sound pronounciation by referring to the formation and by helping of the recall of the visual picture with the aid of hand gestures. We also register the sound in writing and the patient may be able to pronounce it alone.

The recall of the consonants and their differentiation from other sounds of the language is started when the four vowels are at a good level of recognition. The sequence of recall of consonants coincides with that of the sounds in the natural development of a child /5/, or with the universal language rule.

On sound-formation level we do not help the patients relearn the sounds of speech, only the way how to recall them. This is proved by the fact that when we make the patient differentiate more and more sounds and insert them in the appropriate sound-sequence, the elements missing from the subject's sound store /i.e. the empty places in it/, appear by themselves. By recognising the first consonant the patient is able to construct a sequence of sounds-and-letters corresponding to a meaningful word from his or her limited number of sounds and letter store. This prodedure is facilitated by adding a new sound and its corresponding letter, by changing sounds, changing syllables, and, in the end, by changing words. Later we broaden the circle of words to be recalled, applying antonyms, synonyms, semantic field, partial and total connections, functional connections, and situational texts. Consequently the actual visualization of the words cannot take place without content. The objects in the situational texts and functional connections obtain sense on the basis of the patient's knowledge in connection with acting.

We consider it appropriate if the patient's performance is supported by analogue-sentences as soon as possible. On the basis of surface structure characteristics /word-order, suffixes, etc./ these analogue-sentences make empirical generalization possible without the necessity of denominating grammar categories. If the patient can use a new sentence structure by himself or herself, we support this initiative. These are accompanied by real grammar exercises known well from special therapies elaborated for agrammatism /4/.

When practising time-order the patient connects sequences of events. He or she is helped in it by pictures illustrating the sequence of events, and by oral or written questions. Those perception strategies are of importance which have an effect on the text structuring level and enable the subject to adjust the next sentence to the previous one /10/. On this level we reinforce the ability of reading, writing, and retelling fluently the content. In the translation of the fables by La Fontain there are only few redundant elements, so almost every word represents a new information for the patient. This results in a considerable limitation of guessing based on stereotypy of the subject. Therefore the reading strategy can be established on the grounds of sound and letter correspondence, of rhythmical reading of syllables, or of comprehension and expression of whole words.

Finally we may state that the syndrome-specific exercises follow one another with a short time-leg. At the same time we know that the analysing and synthesizing operations work simultanously in the speech process at all levels of the language system.

In the finishing, consolidating phase the patients take part in group therapy, where after a literary work is read they are equal partners in communication and changing information.

References

1. Crystal,D.: (1980) Introduction to Language Pathology. Edward Arnold, London, pp.55-122.

2. Darley,F.L.: (1982) Aphasia. W.B. Saunders Company Press. Philadelphia, pp.1-55.

3. Hegyi,Á.: (1988) Neurolingvisztika a logopédiában. Művelődési Minisztérium Gyógypedagógiai Továbbképző Könyvtára 28, pp.152-159.

4. Jaeger,M.: (1986) Sprache-Stimme-Gehör. X.4: 147-152.

5. Jakobson,R.: (1969) Hang-Jel-Vers. Gondolat Kiadó, Budapest, pp.74-89.

6. Kassai,I.: (1983) A fonéma realitása a korai gyermeknyelvben. Magyar Nyelvőr 4: 467-468.

7. Lang,Ch., von Stockert, Th.R.: (1986) Zum gegenwartigen Stand der Aphasietherapie. Fortschr. Neurol.Psychiat. 54:119-137.

8. Lebrun,Y.: (1983) Cerebral dominance for language. Folia Phoniatrica 35:13-39.

9. Merdian,G.: (1984) Semantische Störungen bei Aphasikern. Europaische Hochschulschriften P.Lang, Frankfurt am Main.

10. Pléh,Cs.: (1981) Különböző szórendű mondatok értelmezése és a dichotikus hallási aszimmetriák 3-6 éves gyermekeknél. Pszichológia 1:365-393.

11. Poek,K.: (1982) Klinische Neuropsychologie. Thieme, Stuttgart.

12. Sági,J.: (1985) A jobb agyfélteke szerepe a beszédmegértés helyreállításában afáziás betegeknél. Pszichológia 5: 435-450.

13. Soos,I., Török,G.: (1976) A beszédhang megkülönböztető képesség és készség kérdésköréhez. Magyar Nyelvőr 3: 309-316.

14. Török,G., Vekerdi,I.: (1981) A beszédhangok tudatosításának kifejlesztése óvodában . In: Az óvodai anyanyelvi nevelés továbbfejlesztése, Kecskemét, pp.102-117.

15. Vekerdi,I.: (1981) Játékos fonémahallás fejlesztés óvodában . Gyógypedagógiai Szemle 3: 199-202.

16. Vekerdi,Zs.: (1981) Óvodáskori beszédhallás fejlesztés. Gyógypedagógiai Szemle 3: 202-206.

17. Weigl,I.: (1979) Neuropsychologische und psycholinguistische Grundlagen eines Programms zur Rehabilitierung aphasischer Störungen. In: Peuser,G. /Hrsg./: Studien zur Sprachtherapie. Fink, München, pp.491-514.

18. Wepman,J.M. et al.: (1960) Studies in aphasia: background and theoretical formulations. J.Speech Hearing Disord. 25: 223-332.

EXAMINATION OF SPEECH AND LANGUAGE
- FROM SPEECH-THERAPIST'S POINT OF VIEW

Ágnes JUHÁSZ and Erzsébet T. BITTERA
Institute of Speech Improvement
Special Teacher's Training College, Budapest, Hungary

In the case of deficit and delay in the development of speech and language the goal of speech-language therapy is to develop and train the expressive, communicative, and nominative function of language, in order to make the child be able to communicate with speech in accordance with his or her age and cognitive level. The speech and language therapy can begin from 2 or 3 years and it can continue for years.

To set the diagnosis it is necessary to examine the child thouroughly because above the deficit in the spoken language we discover partial deficiency of abilities or universal backwardness. On table 1 you can see what sort of examinations are to be done before or during the therapy.

Now, we take out the examination of speech and language development. We put this examination together in the Institute of Speech Improvement -- the practising school of Special Teacher's Training College -- on the basis of special literature and therapeutic experiences with children from 2 to 6 who have grave deficits in their speech and language development.

We can characterize our examination as a process because one meeting is rarely sufficient to devise a sure diagnosis. The experiences during the therapy can modify the diagnosis and in the same way they can modify the content and methods of the therapy. The tasks are built on each other but their order is not severely definite and they aren't age-specific.

After a shorter or longer time of being together with the child we decide which task to start with and how to widen the sphere of tasks. As any of the components of speech can have a deficit a delay or a disorder, we selected the tasks in a way that we can get an aspect from the whole language system of children. We examine verbal comprehension, language reception, verbal expression, language production, semantical and grammatical components, the use of language (pragmatical component) and the manifestation of cognitive functions in speech. We do these within the framework of playing activity which is the basic activity of child.

Our primary goal is to gain useful data for the therapy. (Parts of our examination for speech and language performance are in table 2.) You can see some parts from the examination on videotape and the forms for the tasks.

SYSTEM OF EXAMINATIONS

I.

Statement of the problem
Pregnancy and birth
Development during the first year
Medical history
Family history
Development of verbal comprehension, verbal expression, language production and communicative skills
Development of motion and lateral dominance
Educational history

II.

Speech and language
State and function of speech organs
State of big motion-development, lateral dominance
Knowledge of body parts, orientation in space
Skills of hand motion and drawing
Visual-motor coordination
Hearing perception, hearing attention, hearing discrimination
Visual perception, visual memory
Playing
Cognition
Behaviour

III.

Medical examination (neuropsychology, audiology, phoniatry, etc.)

IV.

Psychological, special pedagogical examination

Table 1.

STATE OF DEVELOPMENT OF SPEECH AND LANGUAGE

1. Nouns
2. Verbs
3. Comprehension of sentences
4. Suffixes of nouns (plural, object, adverbial)
5. Possessive suffixes and pronouns
6. Short term verbal memory
7. Motion and speech integration in sequence
8. Spontaneous specch

Table 2.

SPEECH UNDERSTANDING AND DYSLEXIA: EXPERIMENTAL EVIDENCE

Mária LACZKÓ
Linguistics Institute of the Hungarian Academy of Sciences
Budapest, Hungary

Introduction

It is no exaggeration to claim that the number of poor readers in our school is on the increase. More and more schoolchildren tend to have difficulties in understanding what they read or giving expression to what they think. The proportion of such children is rather high in all primary schools and in most secondary schools as well. This observation is verified by numerical data gathered from various investigations. In an international survey conducted in 1970, testing the reading/understanding level of 10, 11, and 18-year-old subjects from over 20 countries, Hungarian children came out last but one (1). In 1980, another test was administered to 14-year-old pupils. The startling results were as follows: 40% of the participants understood what they read and were able to draw conclusions; another 40 % of them understood each sentence but were unable to draw any conclusions; and 20% exhibited partial comprehension or none at all. A greet deal of tape recordings (made for the Survey of Spoken Hungarian, Linguistics Institute of HAS) bear evidence of subjects of almost all kinds of social background reading haltingly, with a lot of hesitations and multiple false starts.

These phenomena characterise, in extreme cases, the reading and writing inability called dyslexia. Researchers have proposed several explanations for the "syndrome" of dyslexia. Examining the supposed causes of dyslexia, concerning language ability, the question suggests itself: Can deficiencies of hearing and/or the speech understanding/perception process be responsible for reading/writing defects, and if they can, to what extent?

Method and material

Series of experiments with 33 twelve-year-old pupils from Budapest primary schools have been carried out. In these experiments, the subjects were children whose school performance was permanently poor in all subjects or who were close to fail in Hungarian grammar and in their second language, Russian. In all experiments a control group was used, also from Budapest schools, without any selection. The series of experiments consisted of 4 parts. 1. Hearing-perception-understanding tests with the G-O-H method. (Participants heard 10 artifically-produced monosyllabic Hungarian words separately administered to each ear. They had to repeat these words immediately.) 2. In a sentence understanding task they had to repeat 10 well-formed Hungarian sentences masked by white noise. 3. Judgement of comprehension of a coherent text. (Participants heard about 1.5 minute detail of a 19th century, Hungarian novelist's novel and

they have given answers to 10 questions in writing, controlling the comprehension of the text.) 4. Verbal and visual memory tests were also carried out with the word lists consisting of mono-, di-, and trysillabic noncompounds (with the exception of the compund 'snowman').

Results and discussion

The G-O-H screening of hearing revealed that 8 children, 24.3% (!) of the sample, had 'hearing defects'. Two of them had problems in both ears, four in the right ear, and two in the left ear only; the loss was between 15 and 25 dB. Audiological examination was carried out in two cases (see more in detail:2) Among them the G-O-H data for W.Z. (age 12) were corroborated by the clinical examination which verified the right ear hearing loss. The medical certificate said W.Z. had had a radical operation in the right ear for cholesteatoma filling up aditus and barrel that had damaged the auditory ossicles as well. His sentence understanding score was 10%; the number of correctly identified words in the sentences was also low, 25.8%, and the comprehension of coherent text was rather poor, too. It appears that his hearing loss contributed to his poor result in reading, mother tongue skills, and second language learning. This case should draw attention to the importance of permanent screening of hearing and speech-perception process.

The average "speech perception" performance of the test group was much lower than the required performance of the control group of the same age (only 70% correct responses); for several individuals even lower, 50-65%. This performance meets the requirements of 5-year-old children. Hence this may be on of the causes underlying reading comprehension problems that will in turn show up in poor school performance, in particular mother tongue skills and second language learning.

That assumption is further confirmed by the identification scores of the masked sentences (see Table 1).

Table 1

Groups tested	Correct answers	
	masked sentences	words ocurring in those sentences
5-year-olds	36.2%	57.6%
10-year-olds	46.4%	68.4%
14-year-olds	52 %	78.1%
12-year-olds with suspected dyslexia	18.4%	40.9%

These results, when compared to scores of the control group, correspond to the required level of 4-year old children. Consequently, text processing proves difficult for the test group. These children are simply unable to understand the content of most texts presented to them. All these results prove their inability to gather the essence of what they read of to draw conclusions from it; all they understand in most cases are

fragments or single words. It is also supported by their performance in the text comprehension task that is intended to test the coordinated operations both of the complete speech understanding/perception process, and that of its different levels. The average speech perception level of the test group is 34%, and it is about 20% worse than that of the control group.

These children answered the various questions very poorly, too (Table 2).

Table 2

Questions	Results	
	suspected dyslexia group	control group
for a place	15.4%	21.8%
for an object	69.6%	87 %
for a figure	35 %	56.5%
for a colour	46 %	78,4%
for a number	23 %	43.5%
for a promise	57.7%	70 %
for a comment	7.6%	61 %
for a word meaning	0 %	8.8%
for a place	57.7%	56.5%
for a reason	27 %	56.5%
average	34 %	54 %

Some of the questions are so-called "key issues" that have a main part in the comprehension of the whole text (e.g. questions 2 6 9 and 10). The other questions inquire about details (1 5 7). It seems that poor readers are able to give an acceptable answer only to "key issues", and they can't answer questions about the details, especially if they have to answer questions that don't have any coherence with the descriptions of the given "action" (see question number 7). On the other hand, the control group understands the "key issues" and the details about to the same extent as expected. It appears that the speech understanding process of dyslexia suspects (as opposed to the speech understanding process of the normal pupils) is similar to that of adults in that it is global. These pupils are not able to analyse problems that need details. Their poor school performance can be explained by these results. It might also be suggested that possible dyslexics' low scores are primarily due to the inadequate functioning of the acoustic, phonetic, and phonological levels of the speech understanding/perception process, as well as their impaired capacity of verbal shortterm memory.

The set of phenomena involved further includes the functioning of short-term memory. The group of suspected dyslexics achieved an average of 5 words on the verbal memory task, on the visual memory task their achievement was a little better, (the same as that of the control group). That again indicates problems of storage and recall. This point can be illustrated with M.K.'s test results as shown in Table 3.

Table 3

Tests	Results
G-O-H results	35 %
Correct answers for masked sentences	0 %
Correct answers for words occurring in sentences	9.7%
Comprehension of coherent text	20 %
Verbal memory	3 words
Visual memory	4 pictures
School performance	poor
Hungarian grammar	2*
Second language (Russian)	1-2

*Note scale in Hungary: 1(worst) - 5 (best)

The term dyslexia is used to cover reading and writing disturbances of different origin, extent and types. The present experiment confirms and demonstrates that serious difficulties of speech perception and understanding tend to underlie the poor school performance and actual reading defects of dyslexics or suspected dyslexics. These children with their poor results in speech perception and understanding - which was a surprise even for their teachers - could make their way as far as the sixth grade and will soon have to decide what career they want to choose. Consequently, the importance of continually checking children's speech perception and understanding, indeed the need for a well-organized system of regular screening, cannot be over-emphasized.

References

ADAMIKNÉ JÁSZÓ Anna: A magyar olvasástanítás története a kezdetektől napjainkig. Budapest /Kézirat/

Mária LACZKÓ: An experiment on the relation between speech understanding and dyslexia. Hungarian Papers in Phonetics 19. 1988, 82-91.

TACTILE RECODING OF PHONOLOGICAL FEATURES IN A SYSTEM FOR ELECTROCUTANEOUS SUBSTITUTION OF SPEECH FOR THE DEAF

Hans Georg PIROTH
Institut für Phonetik und sprachliche Kommunikation
Universität München, F.R.G.

Introduction

In previous investigations the concept of 'quasiarticulatory tactile speech substitution' was developed (1,3). This 'speech-to-skin' transformation method is based on the assumption that one main component of speech perception consists in the reconstruction of articulatory gestures out of the auditorily analysed acoustic signal. In its actual developmental phase the system only contains the construction of tactile syllable equivalents from an articulatory description of the syllable. This quasiarticulatory coding method starts with a set of coding rules for the transformation of articulatory phonological features into features of dynamic tactile patterns.

Preliminary experiments have shown that the tactile features are well recognizable if they are presented in simple patterns in standard discrimination and identification tests and do not exceed the information transmission capacity of the skin senses which is limited by spatial and temporal masking in a way different from the auditory system (1).

The coding rules determine the number of tactile impulses and the place of stimulation in electrocutaneous sequences of short bipolar pulse trains produced by the 16-channel stimulation device SEHR-2 as described by Tillmann and Piroth (4). To preserve a general geometrical similarity to the vocal tract the left forearm is used as stimulation area and the 16 electrode pairs are spread nearly equidistantly around it between wrist and elbow, building up either four lines or, if regarding the arrangement from another point of view, of four rings of four electrode pairs each as shown in Fig. 1.

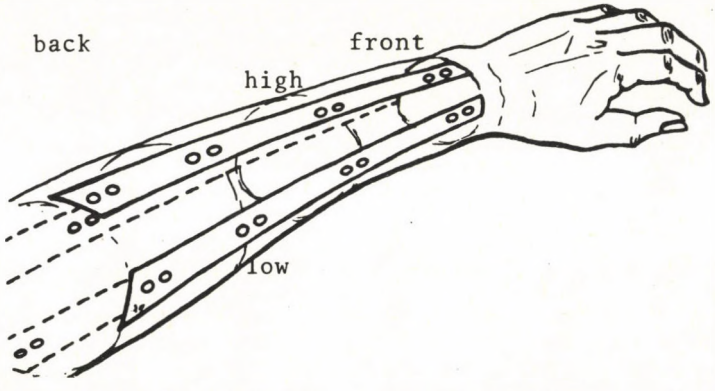


Figure 1
Electrode Arrangement

The General Feature Transformation Rules

1. Vowels and Consonants

Vowels are coded by dynamic tactile patterns moving longitudinally along the arm, consonants by patterns surrounding the arm.

2. Vowels: Tongue Height and Tongue Position

Front vowels are coded by patterns moving along the arm near the wrist, back vowels near the elbow. High vowels move along the dorsal side, low vowels along the volar side of the forearm. Intermediate tongue heights and positions are coded at places between the extreme values.

3. The Place Feature in Consonants

Front consonants are coded as patterns surrounding the arm near the wrist, back consonants near the elbow. In this way four places can be transformed. Additional places between the four rings are coded as patterns surrounding the arm at both neighbouring rings nearly synchronously.

4. Consonantal Modes of Articulation: Plosives and Fricatives

Fricatives are coded as moving around the arm with a constant velocity, plosives as oscillating between neighbouring electrodes.

5. Consonantal Modes of Articulation: Fortis and Lenis

Fortis patterns are coded as moving faster than lenis patterns. In fortis fricatives more impulses are delivered to create the circumferent pattern than in a lenis pattern, but the intervals are shorter to ensure an equal overall duration of fortis and lenis patterns. In fricatives the circumferent pattern moves faster or slower, in plosive patterns the circumferent impulse sequence (resembling the burst) is identical for fortis and lenis patterns, and the velocity difference is encoded into the first half of the following vowel pattern (resembling the transition).

Other features to code phonemic categories such as liprounding, liquids, nasals, and trills have not yet been included at the present stage in the development of the coding method, but several tactile features can be assumed to be suitable transmission of the necessary amount of additional information, e.g. variations of stimulus intensity, impulse form or the direction of movement in the pattern. In addition, all patterns are formed in a way that allows temporal lengthening and shortening to enable the transmission of durational variations in a later phase.

General Experimental Procedure

For each experiment a limited set of six or eight tactile syllable equivalents (one variable phoneme in a fixed context) was selected. The experiments were designed as identification tests with eight or six repetitions of each pattern. Preceding each test-run stimulus intensity was adjusted by the subjects to ensure that stimulus magnitude is within the appropriate range for elctrocutaneous stimulation. After that calibration procedure, but before the identification test was run, the set of items was presented to the subjects a few times in a systematic order to make them familiar with the range of diversity in the items under test.

Each subject was tested singly and had to give his response immediately after a single presentation of the stimulus. When spontaneous

identification was tested the next item was presented after a fixed interval. When training was included subjects received a feedback after each response and had to call the following stimulus by pressing a knob. In these cases the tests were repeated several times with a pause of two or three days between two successive sessions and the subjects were informed of their results after finishing a session.

The Experiments

Each experiment was designed to test one or two of the feature transformation rules. The purpose of Exps. 1 and 2 was to determine the spontaneous identification rate of tactile vowels and tactile consonantal places of articulation. In Exps. 3 and 4 training was involved (cf. Piroth (2) and (3)). Tab. 1 gives the number of subjects, the set of items and the transformation rules under test, the spontaneous identification rate (first session) and the results of training (last session) for each experiment.

Table 1

Exp.	N		Inventory	Context	Rule	1th Session	5th Session
1	5		/i,e,ɛ,a,o,u/	/hV,Vh/	2	65.8 %	-
2	5	(a)	/f,ʃ,x,h/	/Cə,əC/	3	42.2 %	-
		(b)	/s,ç/	/Cə,əC/		30.6 %	-
3	4	(a)	/f,ʃ,x,h/	/Cə/	4,5	47.2 %	65.7 %
		(b)	/s,ç,z,j/	/Cə/		43.7 %	57.8 %
4	4		/p,t,k,b,d,g/	/Cə/	4,5	52.5 %	61.8 %

Concerning Exps. 2 and 3 the set of items has been divided into one-ring patterns (a) and two-ring patterns (b) according to rule 3. The spontaneous identification of vowels yields 65.8 %. Spontaneous recognition of consonants is poorer, but reaches an average of more than 60 % after five sessions.

In analysing the form of the patterns presented for identification several tactile stimulus parameter categories can be distinguished:

1. 2-dimensional local coding (representing tongue height and position in vowels)
2. 1-dimensional local coding (representing consonantal places of articulation)
3. Temporal simultaneity coding (representing additional places of articulation)
4. Linear temporal coding (representing the fortis-lenis feature)
5. Nonlinear temporal coding (representing discontinuities as in plosives)

The results show a clear advantage of local coding: patterns distinct in Dimension 1 can be recognized almost instantaneously (Exp. 1), distinct in Dimension 2 after brief training (Exps. 2a and 3a). The two-ring patterns (Dimension 3) are hardly to be distinguished from one-ring patterns in the beginning (Exp. 2b) and their recognizability remains lower than for one-ring patterns until the last session (Exp. 3b). This might be caused by spatial masking phenomena that become active when the impulse trains are presented nearly synchronously. The temporal coding of Dimensions 4 and 5 (the distinction between plosive and fricative patterns

and the tactile equivalent of the fortis–lenis feature) show a learning effect with a small growth over five sessions (Exps. 3 and 4). Although the results exceed 60 %, it must be taken into account that the subjects started at a relatively high level because of their greater experience and familiarity with tactile speech stimuli after their participation in Exps. 1 and 2.

For the incorporation of other features into the coding method a sixth dimension of tactile parameters has been under consideration:

6. Intensity coding (as an additional parameter)

In preliminary tests using stimulus intensity as a coding parameter only a few levels could be discriminated in similar patterns. But it can be supposed that Dimension 6 can be used to code at least an additional binary feature.

On the whole, the results are encouraging since the system seems to incorporate enough degrees of freedom to create as many distinctions as are necessary to transform a sufficient number of phonological features into the tactile mode to carry all distinctive linguistic information. If it is possible to advance to this level an attempt to include phonetic variations becomes realistic. In this case the aim of quasiarticulatory tactile speech transmission will be the coding of running speech including durational variations, phonetically reduced forms and suprasegmental properties to enable not only the transformation of distinctive phonological features but a real–time mapping of articulatory gestures to the skin.

References

1. Piroth, H.G.: Elektrokutane Silbenerkennung mit quasi–artikulatorisch kodierten komplexen zeitlich–räumlich strukturierten Reizmustern. Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation 22. München, 1985.
2. Piroth, H.G.: Electrocutaneous syllable recognition using quasiarticulatory coding of stimulus patterns. Journal of the Acoustical Society of America 79, Suppl. 1. 1986, S73.
3. Piroth, H.G.: Incorporation of the fortis–lenis feature in a quasiarticulatory system of tactile speech synthesis by adding temporal variations. Proceedings of the 11th International Congress of Phonetic Sciences, Vol. 1. Tallinn, 1987, 369––372.
4. Tillmann, H.G.––Piroth, H.G.: The electrocutaneous stimulation system SEHR and the perceivability of tactile syllables. In Laver, J.––Jack, M.A.: Proceedings of the European Conference on Speech Technology, Vol. 1. Edinburgh, 1987, 423––426.

IMPROVED CONSONANT VOICING PERCEPTION BY HEARING-IMPAIRED LISTENERS FROM SPEECH CUE ENHANCEMENT

Sally REVOILE and Lisa HOLDEN-PITT
Center for Auditory and Speech Sciences
Gallaudet University, Washington, D.C., United States

Introduction

In anticipation of future hearing aids that will exaggerate speech acoustic cues, we have been studying the effects of cue enhancement on consonant recognition by hearing-impaired listeners. Of particular interest are differences among the hearing-impaired in their ability to use enhanced acoustic cues, because the future aids will likely be programmable for users individually. Thus, it is important to examine whether cue enhancements differ in their effectiveness as a function of listeners' audiological profiles &/or other types of auditory capacities. Our initial investigation of such relations concerned enhancement of the vowel duration cue for voicing distinctions of final fricatives. We found that use of enhanced vowel duration cues by severely/profoundly hearing impaired listeners was related to a combination of factors involving low-frequency hearing and listening levels, as well as auditory duration discrimination (1). Subsequent work has focused on enhancement of other cues to final consonant voicing distinctions such as amplification of the constriction or release segments of stop (2) or fricative (3) consonants. That work is examined further in this paper through comparison of the effectiveness of enhancements of both stop and fricative consonants for the same group of severely/ profoundly hearing-impaired listeners.

Method

Stimuli. The stop and fricative consonants were contained in /bəɛC/ syllables spoken by a female. Representing each consonant, four different utterances served as the test stimuli. These tokens had been selected from a larger pool of recorded syllables, to differ minimally in vowel duration between voicing cognate utterances. Following A/D conversion (16.67 kHz sampling), durations of phoneme segments within the syllables were measured on waveforms of each utterance; rms intensity was measured across the segment durations. Mean duration and intensity characteristics of the segments from the natural utterances appear in Table 1.

The syllables with enhanced consonants were prepared using the sets of natural utterances. To yield the enhanced /bəɛg, bəɛd/ utterances, the closure murmurs were amplified by 11 dB to approximate the level of the preceding vowels. To obtain the enhanced /bəɛk, bəɛt/ utterances, the final release bursts were

TABLE 1

For 4 utterances each of /bæk, bæt, (etc.)/, means (& standard deviations) of segment durations and intensities. Both the natural and enhanced intensities are shown for the consonant segments.

	Duration in msec			Intensity dB rms: arbitrary ref.				
	Vowel	Stop occlusion	Stop burst /Frication	Vowel	Occlusion murmur		Stop burst /Frication	
					Nat.	Enh.	Nat.	Enh.
/bæk, bæt/	239 (9)	111 (8)	47 (10)	50 (1)	20 (1)		33 (2)	47 (2)
/bæg, bæd/	244 (12)	67 (12)	49 (14)	51 (1)	40 (2)	51 (2)	33 (2)	
/bæf, bæs/	290 (8)		211 (19)	49 (1)			28 (2)	48 (0)
/bæv, bæz/	292 (10)		114 (14)	49 (1)			28 (2)	50 (2)

amplified by 14 dB, the maximum amount possible without peak clipping of the bursts. For enhanced /bæf, bæs/, the frications were low-passed filtered (5kHz) and amplified by 21 dB. Enhanced /bæv, bæz/ were obtained by replacing the /v, z/ frications with iterated pitch periods copied from the end of the vowel in each utterance. These pitch periods contained some F0, as well as consonant frication noise. The pitch periods were iterated to yield frication durations that nominally matched those of the deleted frications. The segments were band-pass filtered (0.25 to 1kHz) to reduce the presence of the vowel F0 and intensified by 18 dB.

Procedure. Consonant recognition was assessed via syllable-identification trials requiring a response from an answer-set that showed the syllables under test. The syllables with stops and fricatives were tested in separate blocks, as were the utterances with natural or enhanced consonants. In each block, each utterance was presented three times, yielding a 48-trial block (4 consonants X 4 utterances X 3 presentations). The stimuli were presented at each listeners' most comfortable level to the ear with the better hearing thresholds.

Among $\sim$20 one-hour listening sessions, the experiment began with practice syllable-identification trials using non-test stimuli, followed by baseline assessment of consonant recognition for the natural and enhanced stop and fricative stimuli, and subsequently, training and further assessment of consonant recognition. In each session that included training, consonant recognition was tested immediately before and after the presentation of several blocks of training trials. These trials enabled listeners to compare words in pairs that differed only for final consonant voicing (1). A maximum of 4 training sessions was devoted to each set of the enhanced stop

or fricative stimuli, and subsequently, about 2 sessions to each of the syllable sets of natural stops and fricatives. Fewer sessions were used for the natural syllables because they yielded generally poorer performances than the syllables with enhanced consonants.

Subjects. Severely/profoundly hearing-impaired undergraduates (N=18) at Gallaudet participated as listeners. Their mean hearing threshold across .5, 1, and 2kHz was 93 dB HL, and ranged from 71 to 110 dB HL among listeners.

Results

The listeners' consonant identification was scored according to percent correct voicing recognition, that is, errors in perception of consonant place were ignored. The mean scores in Table 2 suggest that the listeners' ability to distinguish consonant voicing was facilitated by the training, but was especially influenced by the acoustic cue enhancements. Prior to training, the listeners' ability to distinguish voicing was generally poor for the natural as well as the enhanced consonants. The baseline tests for the fricatives when natural and enhanced yielded chance level performance (i.e., <60%) for more than three-fourths of the listeners, while for the stops natural and enhanced, about one-half of the group performed at chance on the baseline tests. After training for the enhanced fricatives, nearly one-half of the listeners continued to show chance performance and consequently, were not presented training for the natural fricatives. (In Table 2, this explains the smaller number of listeners (n=12) with scores for the natural fricatives after training). For the enhanced stops after training, only one listener performed at chance level.

TABLE 2

For 18 or (12) hearing-impaired listeners, mean % correct voicing perception for final consonants with natural or enhanced voicing cues. Scores are shown for baseline tests and tests after training.

	Natural		Enhanced	
	Fricatives	Stops	Fricatives	Stops
Baseline	47	55	50	60
After training	(61)	67	71	81

Among those listeners who performed above chance for the enhanced consonants after training (n = 11) voicing perception was significantly better for the enhanced stops than for the enhanced fricatives [t test for paired samples: $t(10) = -4.1$, $p < .01$]. However, when these listeners' scores for the enhanced consonants after training were compared to their scores for the natural consonants after training, the amount of improvement among the enhanced consonants was similar between the stops and frictives [$t(10) = -.67$, $p = .52$]. Thus, the higher performance seen for the enhanced stops versus fricatives may be related more to the inherent differences in perceptibility of the stops versus fricatives for these listeners, rather than an indication of greater salience

achieved by the stop enhancement in comparison to the fricative enhancement.

Some differences in hearing and syllable presentation levels are apparent in Table 3 between the listeners who performed above versus below chance performance for the enhanced stops and fricatives after

TABLE 3

Mean audiometric thresholds and syllable presentation levels (dB SPL) for listener subgroups who performed above (n=11) or below (n=7) chance level for the enhanced stops and fricatives after training.

	Frequency, kHz							
	.25	.5	1	2	4	syllable level	/æ/vowel/ threshold	/æ/ SL
11 Ss scored above chance	101	100	97	100	108	118	93	25
7 Ss scored below chance	101	101	108	116	125+	114	99	14

training. The 7 listeners who scored below chance had mean tone thresholds that were at least 10 dB poorer at frequencies $\geq$ 1kHz, than listeners who scored above chance. These poorer thresholds could explain the "below-chance" listeners' apparent difficulty in using the amplified /t/ and /k/ bursts and /f/ and /s/ frications, which contain significant energy in the mid and high frequencies.

While the two listener subgroups had similar mean tone thresholds at frequencies below 1kHz, the "below-chance" subgroup heard the syllables at a lower relative level above threshold than the "above-chance" subgroup, as indicated by the far-right column in Table 3. For the "below-chance" subgroup, the amplified closure murmur of the /g/ and /d/ may have been at or near threshold as a result of the low sensation level (SL) at which the syllables were heard by these listeners.

References

1. REVOILE, S., HOLDEN-PITT, L., PICKETT, J., & BRANDT, F.: Speech cue enhancement for the hearing impaired: I. Altered vowel durations for perception of final fricative voicing. Journal of Speech and Hearing Research 29. 1986, 240--255.

2. REVOILE, S., PICKETT, J., HOLDEN-PITT, L., EDWARD, D., AND BRANDT, F.: Speech-cue enhancement for the hearing impaired; II. Amplification of burst/murmur cues for improved perception of final stop voicing. Journal of Rehabilitation Research and Development 24. 1987, 207--215.

3. REVOILE, S., HOLDEN-PITT, L., EDWARD, D., & PICKETT, J.: Speech cue enhancement for the hearing impaired: III. Amplification of frication for improved perception of final fricative voicing. Proceedings of the 11th International Congress of Phonetic Sciences 3. 1987, 332--335.

OUR EXPERIENCES WITH APHASIA-REHABILITATION IN 1986-87 YEARS

M.STEPPER, E.FOGAS-TARNÓCZKY
Dept. of Phoniatrics and E.N.T. of Makó-Hospital,Hungary

In the years 1986-87, 53 cerebrovascularly disordered aphasic patients were tested and treated in the Outpatient Dept. of Phoniatrics of Hospital-Makó. We have no possibilities to carry out an intensive inpatient speech-rehabilitation course for the aphasic-patients, thats' why we make our first informative checking of speech abilities - beside the beds of patients - at the Dept. of Neurology and Dept. of Internal Diseases of Makó-Hospital after having settled the patients somatic conditions, averagely 6 weeks after the onset of their diseases.

The first checking included the following items: spontaneous speech, automatized speech, speech imitation, word-finding, speech comprehension, spontaneous writing, dictation, reading and the checking of the occipito-parietal symptoms. The aim of this informative checking is to determine the initial severity of aphasia and it serves as a basis for the first aphasia diagnosis as follows:

No.	Type of aphasia	Age	Average age	Male	Female
15	Broca	50-86y.	54,8 years	7	8
5	Global	51-86"	71,8 "	4	1
4	Transcort.-mot.	51-86"	67,5 "	2	2
7	Transcort.-sensomot.	50-85"	70 "	4	3
8	Anomic	56-80"	66 "	4	4
3	Central	68-82"	73 "	1	2
7	Transcort.-sens.	58-86"	79 "	3	4
4	Wernicke	59-80"	72 "	1	3
53				26	27

Considering an average age we can conclude that, those patients belonging to the group of expressive and anomic speech disorders were somewhat younger than those belonging to the group of recessive speech disorders.

During the first 3-month period of inpatient status they were practically observed and according to their capacities and rest-symptoms 15-20 min. speech exercises were organized for them (2, 3, 4, 7)

Before the patients were let out from the Hospital, their speech ability was repeatedly checked with the I.Lang test together with a polysensoric examination (checking of seeing, hearing, equilibrium etc.). Pure-tone audiometry and speech-audiometry cannot be evaluated with good security in cases of extended cerebrovascular diseases. It is extremelly difficult to establish contacts in those cases where the patients' aphasia is associated with a loss of hearing. In the group of so

called "fluent" aphasia perception is not reliable enough, paraphasia and neologism make it impossible to evaluate speech audiometry. In the group of "non-fluent" aphasia the results of hearing tests are uncertain even in the case of slow-paced tests because of the patient fatique of attention and uncertain responses.We have no possibilities to carry out objective hearing tests so we decided on, testing the acoustic reflex-threshold which in many cases offers a more reliable information on the disorders of the n.acoustic pathways and its degree than the pure-tone and speech audiometry. The acoustic reflex-threshold can reveal brain-stem disorders as well (1). It is well known from the available literary data, that Thalamus plays an important role in functions like word-finding,repeating,naming objects, writing and reading(5). The acoustic reflex response can reveal the laesions of the n. trigeminus and n. facialis in case of patient with dysarthria (9).From our patients 7 suffered from hearing loss and aphasia and 2 from dysarthria aphasia and hearing loss. After having stabilized their somatic condition their hearing were tested repeatedly - bi-weekly - as with the improvement of the central nerve functions (attention, perception, gnosis) the results of the hearing tests were also improving in some cases.The hearing loss was of sensoneuric character. In case of 4 patients with expressive speech disturbances the hearing aid helped their speech rehabilitation. On the other hand in cases of patients with receptive speech disturbances the hearing aid did not help their communication on the contrary, it meant an obstacle for them.

As the results of repeated checking our patients' speech-ability is as follows:

No.	Type of aphasia	Average age	Revealed	Died	Male	Female
8	Broca	54,8 yrs			4	4
5	Global	71,8 "		2	4	1
8	Transcort.-mot.	67,5 "			5	3
7	Transcort.-sensomot.	70 "			4	3
11	Anomic	60 "	4		4	7
3	Central	73 "			1	2
7	Transcort.-sens.	79 "			3	4
4	Wernicke	72 "			1	3
53			4	2	26	27

Before we let the patients out from the Hospital, the decision had to be made, who was to be regularly rehabilitated afterwards.Following averagely a 3-month observation and treatment period, we did not see any point in the rehabilitation of 11 patients with two-sided extended brain damage,serious case of diabetes with sight problem, old age and bad general condition.

Sometimes the family had no possibility to arrange for the patient to attend the speech rehabilitation twice a week (the family-members were of old age too or had no time to accompany the patient because their jobs).

26 patients speech rehabilitation was begun on three occasions per week, each of them was given an individual treatment for a year, after which two groups were formed with 6-6 members.

14 patients speech ability after one year individual treatment was below the level this group rehabilitation would have required, so they had to be treated on individually.

Following the 2 year rehabilitation-program the aphasic patients were re-checked using the I.Lang test and interviews of the family about their speech abilities. The result can be considered favourable if the patient can communicate without help and difficulties in an unfamiliar environment. It can be considered agreeable if the patient can communicate also without help but with difficulties. It can be considered poor if the patient is not able to communicate without help.

Outcome of two years speech rehabilitation of 53 aphasic patients':

Type of A.	Restored	Improved	Unchanged	Change of type of A.	Died	No
Broca	2	1	5	$4_{Tm}+3_{An}$		15
Global			1		4	5
Tr.-mot.	1				3	4
Tr.-s.m.		4	3			7
Anomic	5	3				8
Central			3			3
Tr.-sens.					7	7
Wernicke			1		3	4
Total:	8	8	13	7	17	53

(Tm= transcortical-motoric, An= anomic, Tr.-mot=transcortical-motoric, Tr.-s.m.=transcortical-sensomotoric,Tr.-sens=transcortical-sensoric)

Discussion

There is a need for information and check-ups to determine the initial severity of aphasia, so as to differentiate between the dementia, a serious degree of apraxia and/or dysarthria and aphasia. But this informal checking does not offer a reliable prognosis for speech-rehabilitation. This hypothesis supported by the quick improvement of the four anomic-aphasic cases and the change of aphasia forms in the cases of the seven Broca aphasic patients turning into four transcortical-motoric and three anomic-aphasic during the first three-month period.

When indicating speech rehabilitation we always have to consider the neurological diagnosis, the basic diseases, the general condition and the age.

The succes of the speech-rehabilitation is also determined by socio-familiar factors. If the family does not want to cooperate it is useless to bring the aphasic patient for the sessions.We had cases, too, where the patients and their family

cooperated with the phoniatricians and with the speech-therapeutist and there was no improvement because of the patient central nerve disorders.So we had 20 cases where no improvement occured.

We had 16 cases where the speech-rehabilitation had a good result and the patients could again adapt themselves to their neighbourhood, they felt themselves again as "human beeings" these were the words used by one of our patient.

References

1. BOSATRA, A. : Pathology of the Nervous Arc of the Acoustic reflexes. Audiology 16: 307-315 (1977)
2. EISENSON, J. : Language Rehabilitation of Aphasic Adults Folia phoniat. 29: 61-83 (1977)
3. HUBER,W., MAYER,I., KERSCHENSTEINER, M. :Phonematischer Jargon bei Wernicke-Aphasie Folia phoniat. 30: 119-135 (1978)
4. HUNTLEY,R.A., ROTHI,L.J.G.:Treatment of verbal akinesia in a case of transcortical motor aphasia Aphasiology, 2: 55-66 (1988)
5. MAZAUX,J.M.: Troubles du langage au cours des lesions thalamiques. Rev. neurol. 135: 59-64 (1979)
6. SPITZER, H.: Theoretische und praktische Grundlagen zur Behandlung phonematischer Störungen Sprache-Stimme-Gehör 12: 72-76 (1988)
7. VARGHA, M., GERÉB, GY.: Aphasie Therapie. Samml. zwangl. Abhandlungen aus dem Gebiete der Psychiatrie und Neurologie. 18: Veb. G. Fischer Verl. Jena 1959.
8. VERSEGHI, A.: Az afázia rehabilitációjának módszerei. Osmánné Sági J.: Az afáziák neurolongvisztikai alapjai. Tankönyvkiadó, Budapest, 1986. 87-92
9. WALLESCH, C.W.: Lesions of the Basal Ganglia, Thalamus and Deep White Matter. Brain and Lang. 20: 286-304 (1983)

PERCEPTION AND PRODUCTION OF A VOICING CONTRAST IN FRENCH-ENGLISH BILINGUALS

Georges BOULAKIA, Valérie HAZAN
Université de Paris VII, University College London

Introduction

A number of studies have examined how bilingual speakers perceive and produce contrasts which are marked differently in the two languages in which they are fluent. The voicing contrast in initial plosives is most commonly used for these investigations. The main cue to this contrast is Voice Onset Time (VOT), or the duration between the burst release and the onset of voicing. The contrast in French is between a lead versus short-lag VOT , and in English between a short- versus long-lag VOT. A secondary cue to the contrast in English is the onset frequency of the first formant . F1 onset is not contrastive between stimuli with lead and short-lag VOT. Some experiments with French-English or Spanish-English bilinguals (Caramazza et al.,1973; Williams, 1977) have found no evidence of code-switching between language sets in bilinguals, with the mean phoneme boundary falling between the boundaries obtained for monolinguals. However, using natural stimuli and a more rigorous procedure, Elman (1977) showed evidence of code-switching in strong bilinguals. Weaker bilinguals tended to use the same boundary in both language sets.

Preliminary results are presented here for a study which has focussed on the use, bu bilingual speakers, of the F1 onset cue to the voicing contrast. by bilingual listeners The procedures used in previous studies have been improved in different ways. Firstly, a contrast was chosen,which was formed of meaningful words in both languages (PEN-BEN and PEINE-BENNE). Also, bilinguals based both in Great-Britain and in France were tested in their country of residence, in order to minimise bias due to language of immersion. As the magnitude of phoneme boundary shifts are reduced when synthetic speech is used (Hazan et al., 1987). use was made of computer-edited natural speech.

Stimuli

Test continua were created using digitised natural speech waveforms. In all continua, Voice Onset Time ranged from -40 ms to +40 ms in 10 ms steps. In the first continuum (Pen/VOT), the vowel stem, burst transient and aspiration were taken from the voiceless PEN. In order to create the stimuli with positive VOTs, the aspiration was progressively deleted, in 10 ms slices, following the burst release. In order to create the stimuli with negative VOT values, the prevoiced portion was edited out of a voiced BENNE and appended to the front of the burst release. The prevoicing portion was then progressively cut back from its onset in 10 ms steps in order to create the stimuli with -40 to -10 ms VOT. The continuum thus produced had appropriate VOT values at the extremes of the range, but the spectral characteristics at vowel onset cued voicelessness throughout the stimulus range. In the second continuum (Ben/VOT), the vowel stem from BENNE was used. The same technique as described above was used to obtain the VOT continuum . In this continuum, though, the spectral characteristics at vowel onset cued voicedness throughout the stimulus range.

In order to create the "French" and "English" test conditions, each of the stimuli described above was preceded either by a French ("répète") or English ("repeat") precursor. For each condition, an identification test tape was produced by randomising and recording ten tokens of each stimulus.

Subjects

Four groups of listeners were tested: 8 bilinguals living in London, 11 bilinguals living in Paris, 10 British monolinguals and 15 French monolinguals. All listeners filled in a questionnaire which was then used to define their degree of bilingualism. Bilinguals are

defined as "strong" if they learnt both languages before the age of five, and "mid" if they learnt their second language later, but had spent at least three years in a country in which their second language was spoken. Results are included here for two sets of strong bilinguals (2 subjects with English-bias and 4 with French-bias), and two sets of mid bilinguals (3 with English-bias and 7 with French-bias).

Test procedure

Testing was carried out over two one-hour sessions on separate days. At each session, only one language was used. The session started with a speech recording of "accent-revealing" sentences and of minimal pairs in the language being tested.The subjects also conversed with the bilingual experimenter in the appropriate language. The subjects then listened to two test tapes containing the pen/VOT and ben/VOT test conditions. Stimuli were presented free-field at a comfortable listening level.

Results of perceptual tests

Results obtained by individual listeners were grouped according to their degree of bilingualism and language-bias. Maximum likelihood estimates were applied to the identification functions obtained in order to get a measure of phoneme boundary. The values obtained are presented in Table 1. The mean identification functions obtained are presented in Figure 1.

Table 1: Mean phoneme boundary in milliseconds (with standard deviations)

	Ben/VOT E		Ben/VOT F		Pen/VOT E		Pen/VOT F	
Strong (Fren-bias) bil.	+9.7	(1.1)	+7.0	(1.1)	-25.3	(3.0)	-30.1	(2.8)
Strong (Eng-bias) bil.	+9.0	(1.3)	+4.9	(1.4)	-18.5	(5.5)	-28.7	(4.0)
Mid (Fren-bias) bil.	+11.3	(0.8)	+9.2	(0.7)	-7.1	(1.1)	-10.9	(1.0)
Mid (Eng-bias) bil.	+7.3	(1.3)	+15.7	(2.0)	-42.6	(5.8)	-31.6	(3.6)
Monolingual French			+1.4	(0.6)			-19.8	(0.7)
Monolingual English	+18.8	(1.0)			-17.4	(1.3)		

1. Monolingual English listeners

There is a clear shift in boundary between the two test conditions which differ in vowel onset characteristics. These means hide two types of perceptual behaviour in listeners. Some showed merely a shift in phoneme boundary between the two test conditions while two listeners perceived most stimuli as voiced in the ben/VOT condition, whatever the VOT value, and all stimuli as voiceless in the pen/VOT condition.

2. Monolingual French listeners

As F1 onset is not functioning as a cue to the voicing contrast in French, it would not be expected that differences in F1 onset characteristics should strongly affect monolingual French listeners. The mean identification functions do show that endpoints in both conditions are labelled appropriately, whatever the status of F1 onset. However, there was a clear phoneme boundary shift between the two conditions, showing that the listeners were perceptually affected by the absence of a rising F1 onset,and tend to perceive more stimuli as voiceless, even in the presence of small amounts of prevoicing, if the vowel onset characteristics were inappropriate

3. Bilingual listeners

Averages boundaries obtained by bilingual listeners for the ben/VOT condition appear to be intermediate to those obtained by both groups of monolinguals. There is however a small shift in boundary, within all groups, between results obtained for the English and French conditions, which may be evidence of code-switching. Results are more variable for the pen/VOT condition.An examination of the mean identification functions obtained reveal that for tokens with prevoicing, labelling for all bilingual groups except the mid (French-bias)

bilinguals is similar to that obtained by English monolinguals. Some listeners in all these groups therefore exhibit a strong sensitivity to F1 onset characteristics.

Results of production measurements

VOT measurements (from burst release to the onset of vocal fold vibration) were made from the digitised waveforms on a mini-computer. An average was taken of five measurements for each speaker. Averages are presented in Table 2.

Table 2 : VOT measurements for English and French /pen/ and /ben/ in milliseconds.

	Ben E	Pen E	Benne F	Peine F
Strong (Fren-bias) bil.	-45 (s.d. 53)	+52 (s.d. 16)	-89 (s.d. 59)	+11 (s.d. 2)
Strong (Eng-bias) bil.	-78 (s.d. 33)	+67 (s.d. 3)	-138(s.d. 28)	+9 (s.d. 1)
Mid (Fren-bias) bil.	-65 (s.d. 48)	+50 (s.d. 23)	-116 (s.d.32)	+12 (s.d. 4)
Mid (Eng-bias) bil.	-53 (s.d. 45)	+64 (s.d. 16)	-130 (s.d. 23)	+15 (s.d. 4)
Monolingual English	+9 (s.d. 1)	+72 (s.d. 19)		
Monolingual French			-109 (s.d. 23)	+10 (s.d. 3)

All groups of bilingual listeners have a tendency to prevoice English voiced plosives. The extent of prevoicing was however not as great as that obtained for productions of French voiced plosives. The large standard deviation measures obtained reveal the extent of inter-subject variability in all groups. Intra-subject variability is also found, with some bilinguals producing some /b/ segments prevoiced, and others without prevoicing. A closer examination of the speech waveforms obtained also reveals qualitative differences in the prevoicing obtained, which is of very weak intensity in some bilingual speakers. Bilinguals are producing long-lag VOTs for English voiceless /p/, but generally of shorter durations than those produced by English monolinguals.

Discussion

In the perceptual tests, small shifts in phoneme boundary are found between the French and English test conditions, but there is little evidence of significant code-switching. There is however evidence of difference in behaviour between groups as regards the use of the F1 onset characteristics, which form a secondary cue to the voicing contrast in English. Some English monolinguals, English-bias bilinguals and strong French-bias bilinguals who were exposed early to their second language seem strongly influenced in their labelling of the stimuli in both language sets , to the extent of being quite unable to label stimuli as voiced, even in the presence of prevoicing, if the vowel onset characteristics are cueing a positive VOT. None of the French monolinguals and French-bias mid bilinguals showed such drastic effects of F1 onset characteristics. To all,prevoicing of at least 30 ms duration was an overwhelming cue to voicing, despite inappropriate vowel onset characteristics. At VOT values of -20 ms to +20 ms however, the decision on the voicing characteristics of the token does seem to be influenced by the vowel onset characteristics.

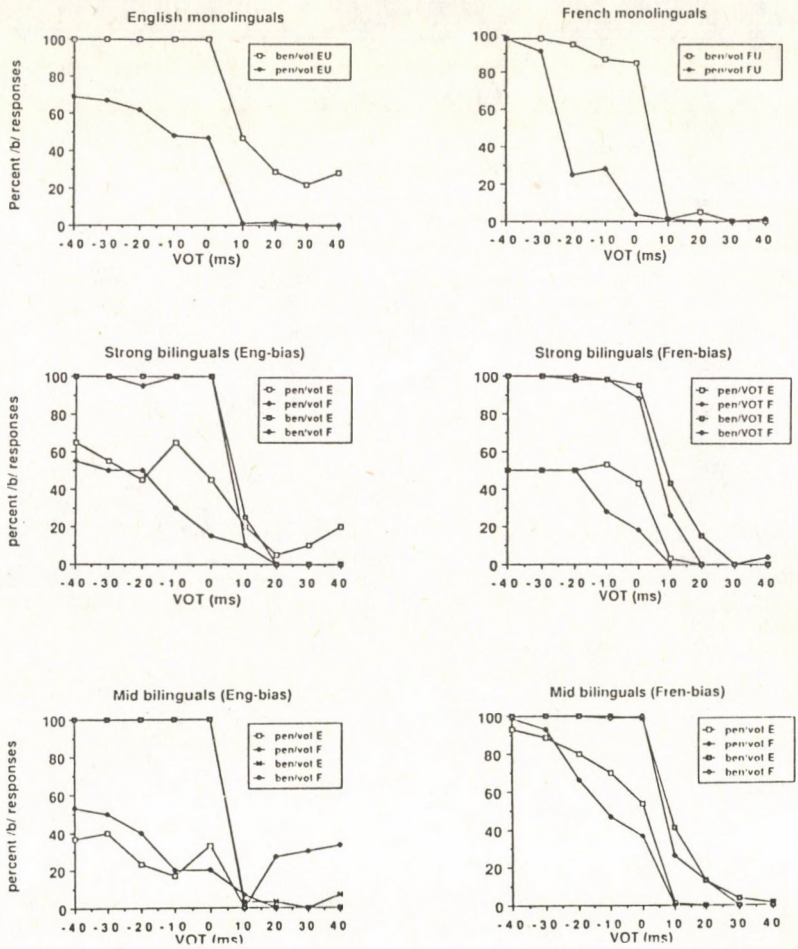
Acknowledgements

We wish to thank Mr P Shaw-Latimer (Head of the English Section of the Lycée International de St Germain-en-Laye) and all subjects for their help and cooperation.

References

1. CARAMAZZA,A., YENI-KOMSHIAN,G., ZURIF E. & CARBONE E. (1973) The acquisition of a new phonological contrast: the use of stop consonants in French-English bilinguals. J. Acoust. Soc. Am., 54, 421-428
2. ELMAN,J., DIEHL,R. & BUCHWALD S. (1977) Perceptual switching in bilinguals. J. Acoust. Soc. Am., 62, 971-974
3. HAZAN,V., HOLDEN-PITT,L., REVOILE,S. & EDWARD,D. (1987) Perception of cues to a stop voicing contrast by normal-hearing and hearing-impaired children. Proceedings of XIth Int. Congress of Phonetic Sciences, Tallinn, USSR
4. WILLIAMS, L. (1977) The perception of a stop consonant voicing by Spanish-English bilinguals. Perception and Psychophysics, 21, 289-297

Figure 1: Mean identification functions for groups of monolingual and bilingual listeners



CEREBRAL ASYMMETRY IN SPEECH PROCESSING

TATIANA V.CHERNIGOVSKAYA, INNA A.VARTANIAN
Lab. of Comparative Physiology of Sensory Systems,
I.M.Sechenov Institute of Evolutionary Physiology
and Biochemistry, Acad. of Sci., USSR, Leningrad

Cerebral functional asymmetry is most commonly described as either depending on the type of material (verbal/visuospatial dichotomy) or the type of processing, i.e. cognitive styles (analytic/holistic). A great majority of empirical research both with normal subjects and brain-damaged patients are interpreted within these dichotomies, the underlying idea being s e p a r a t e functioning of cerebral hemispheres in cognitive processing. However, a large body of data reveal that these paradigms do not represent fundamental differences between the hemispheres: it has been established nowadays that the right hemisphere is quite able to process language while visuospatial material is successfully processed by the left hemisphere. In similar way, both hemispheres can use various cognitive strategies depending on a number of factors including individual differences caused by genetically programmed lateralization of cognitive functions as well as those formed as a result of some specific training. Recent data show that predominant left or right hemispheric involvement in information processing is determined by the task factor - either experimental or real - and consequently the necessity of cognitive style choice: analytic for one class of tasks (used mostly by the left hemispheric mechanisms) versus holistic, Gestalt (used mostly by the right hemisphere). Thus, not only qualitatively different information can be processed by either left or right hemispheric structures but the same stimulus can be described by different hemispheric paradigms depending on the purpose. It is also becoming apparent that the level of stimulus analysis required for the performance of a task is a very important factor: not all the stages of perception (including speech processing) imply hemispheric involvement, i.e. higher cortical functions. Unfortunately most research in speech processing mechanisms do not take into account the above mentioned factors. The problem of interrelation and integration of different aspects of speech-hearing functions (especially sensorimotor resolution capacities) has not

been givendue attention yet. The same is true for the problem of speech perceptual categorization during classification and imitation task performance. In spite of the fact that there are numerous research dealing with different aspects of acoustic parameters important for speech and complex non-speech sounds perception in man the problem of influence of central speech production control on auditory function has not been investigated. The focus of the present paper is finding mechanisms responsible for different aspects of acoustic perception and imitation. Experimental procedure. Experiment I. The subjects were I5 normal listeners between the ages of 20-50 years. All subjects were native speakers of Russian and were right-handed. The stimulus sets were CVC syllables made up of natural speech sounds produced by a male Russian-French bilingual. Russian stop consonants /t/ and /k/ and French and Russian vowels were used to construct on a computer and record the set. The resulting tape consisted of 24 trials with 3-sec. interval which permitted subjects to record their responses manually or vocally. The stimuli were presented monaurally to both right and left ear in turn. Reaction time (latent period between stimulus and response) and the type of reactions were automatically registered. A hemisphere was decided to be dominant for the perception if reaction time for the target heard from contralateral ear was shorter.Subjects were instructed to press the key with left or right hands (in different sessions) (a) - right after hearing the stimulus; (b) - after deciding which syllable they heard; (c) - the same together with the stimulus articulation movement simulation (without phonation). All possible combinations of ears and hands were used. In Experiment 2 subjects were asked (a) - to produce syllable /tak/ right after they heard the stimulus (simple vocal response); (b) - to imitate the stimulus most accurately; (c) - to produce the Russian syllable similar to the target one. All the responses were recorded on a tape recorder and the reaction time was automatically registered. All possible combinations of testing conditions were used. Experiment 3. 49 normal subjects between 24 and 36 years were tested. The stimuli were amplitude-impulse-modulated sounds of different durations. Sounds were noise (frequency range 350-3000 Hz), sustained tones (250,800,I000 and 4000 Hz) and liniarly frequency modulated tones with rizing and falling frequency changes (from 400 to 700 and from 700 to 300 Hz). The duration of a sequence of

pulses was 0.08-3.2 sec., impulses being liniarly rizing or falling. The rythm was 5-80 pulses per second (medium - 30 pulses per second). Subjects were asked to classify the stimuli according to two possible perceptual parameters - speech-like quality and moving in space (approaching or moving away). The stimuli were presented monaurally to the left and right ears in quasirandom order. Subjects were instructed to respond manually (left or right in different sessions). The reaction time was automatically registered. Results and Discussion. Experiment I provided evidence of reaction time hierarchy depending on different task types (a,b,c). The first range - around 240 msec is the time needed just to hear the stimulus and to start reacting manually; the second - around 420 msec - to decide which of the stimuli was presented and the third - to simulate articulation movements of the stimulus without phonation. The greatest reaction time was registered when the stimuli were presented to the left ear, while the response was given by the left hand; the least - when the stimuli were presented to the right ear and the response was given by the right hand. It must be noted that though individual reaction times may vary around the measured value the relation between the ranges remains stable. Experiment 2 also shows hierarchy of latent times - in vocal responses. The simple vocal response of signal detection needs 550 msec, the latent time before accurate syllable imitation - around 920 msec, and categorical judgment of the stimulus to refer it to the native language syllable system (whether it was Russian or French) - 1060 msec. It should be mentioned that processing of native versus "foreign" syllables seems to be controlled by different cerebral structures: "foreign" need mostly left hemisphere machanisms - both for imitation and categorization; probably it is caused by the necessity of phonemic coding, while native syllables can involve both (right and left) hemispheres. Linguistic and acoustic competence of the subjects plays significant role in cerebral balance of speech processing. The preliminary data presented above indicate the importance of taking into account (1) individual differences of the subjects - their linguistic skills; speed of reactions, and preferred cognitive styles; (2) different stages of verbal processing that can be revealed in different experimental procedures. As the role of cerebral asymmetry in speech processing can be shown only by exploiting definite hierarchy of auditory tasks, a special investigation of speed

and accuracy of responses in identification and classification of nonspeech acoustic stimuli was undertaken (Experiment 3). It showed three discrete ranges of stimuli durations revealed in classification tasks of amplitude-impulse-modulated targets according to their perceptual parameters: 0.08-0.2 sec.; 0.2-0.6 sec.; 0.6-3.2 sec. The subjects used these ranges to identify the stimulus as hoarse, speech-like (consonant-like with noise carrier and accent-like with tone carrier) or moving in space (approaching with rizing amplitude and moving away - with falling one). It was shown that classification task is being solved with the same time limits irrespective of the stimulus acoustical parameters - rythm of the pulses, duration, carrier frequency, amplitude shifting, the side of stimulation etc. - in the average-latent time was I.5 sec. However, it should be emphasized that using "speech-like" criterion increases by 30 per cent when the signal is being addressed to the right hemisphere, i.e. to the left ear.The findings suggest that classification procedure in the given experiment was based on dealing with individually formed functionally relevant template recognition. Opposite to it, experiments with amplitude changes identification show basic importance of (a) stimulus presentation side and (b) the use of the right versus left hand for the response. The maximum differences were examined in the range of "speech-like" durations revealed in classification experiment. It is of special interest to note that 20 subjects out of 49 turned out to be grouped in two extremes, the remaining arranged in between. The first one is "reciprocal" and is characterized by sharply different latent times depending on the stimulation sides the physical parameters of the stimuli being identical. The latency for the condition "right ear/right hand" may be changed up to 3 times compared to that of "left ear/left hand" in duration range of 0.2-0.6 sec. (other ranges do not reveal such tendency). As this takes place, dominant functions may prove to be either of the left or the right hemisphere irrespective of subjects handedness (47 were right-handed, 2 were ambidextrous). Subjects of the second group - "synaergic" - have approximately the same reaction time irrespective of the stimulation side, hand or presentation order. This leads us to a suggestion that they use not template recognition but amplitude changes in time. The subjects of this group make significantly less mistakes - up to 2.4 times - compared with those of the

first one. Some of them reveal lesser latent times in the condition "right ear/right hand". In this respect our previous research of auditory function in patients with focal temporal lesions may be of interest. Right-side lesions lead to inability to use speech-like criteria as well as to evaluate signal duration and amplitude changes in the range of 0.1-0.88 sec. Left-side lesions lead to even more serious consequenses: patients are not able to evaluate signals even in the range of up to 3 sec. As a conclusion we put forward a suggestion that in central regulation of speech all high level processing of new and complex information seems to be a function of left hemisphere, while familiar and simple information engages both or the right one only. Speech processing, therefore, most probably uses higher levels in interpreting lower levels of perception. Left hemisphere provides for phonemic encoding and structural analysis of acoustical stimuli using short-term memory. Right hemisphere realizes global template recognition. The data demonstrate two types of sensory-motor organization of subjects. It should be also emphasized that perception may be language-specific and depend on individual acoustic and language background.

MODELLING COARTICULATION IN SYNTHESIZED SPANISH LATERAL CONSONANT [l]

Mª Helena FERNANDEZ[1], Juan M. GARRIDO[2] and Carme DE LA MOTA[2]
(1) Universidad de Alicante (Spain)
(2) Universitat Autònoma de Barcelona (Spain).

0. INTRODUCTION

Several works on acoustic analysis of lateral [l] (*e. g.* CHAFCOULOFF [2]) have pointed out that this consonant does not show a high degree of coarticulatory resistance. The most important variations due to coarticulation have been found in F2 frequency, but little is known about its role in the percepcion of this consonant. The importance of the transitions in perception of [l] has been also investigated in some synthesis experiments. O'CONNOR *et al.* [5] and LISKER [4] studied transitions duration and slope as perceptual cues for [l]; AINSWORTH [1] revealed that short duration of F1 transitions is also relevant . However, it has not been paid attention enough to F2 transitions.

The aim of this paper is to study the role of coarticulatory phenomena in lateral + vowel combinations. The intelligibility and naturalness of the synthesized laterals have been assessed using a battery of perceptual tests.

1. EXPERIMENTAL ANALYSIS

1.1 Stimuli production

Several vowel + lateral + vowel combinations have been sinthesized and used as stimuli in a perceptual test. The generation of these sequences has been carried out in the following way:

a) Steady-state of [l]: formants at 475, 1455, 2575 and 3700 Hz; spectral zero at 2100 Hz.

b) [l] - vowel transitions: F2-F4 starting points placed at [l] F2-F4 frequencies. F1 placed at the same frequency than the vowel.

c) Vowel steady-state: formants at standard frequency values for the five Spanish vowels.

The values for duration have been fixed at 100 ms. for vowels, 60 ms. for the lateral consonant and 5 ms. for each of the four transition steps. [l] has been synthesized with a constant overall intensity level, while the intensity level in vowels is variable. We assume, as it has been previously stated (CHAFCOULOFF [3]), that the relationship between the intensity of the three first spectral peaks in [l] clearly determines the quality of the lateral consonant.

1.2 Elaboration and realization of perceptual tests

Three different tests have been designed in order to discuss the effect of coarticulation in the steady-state of [l], in transitions and in both elements.

Test 1: F2 frequencies in [l] are fixed, but there are three different possible levels for F2 frequencies in transitions.

Test 2: Only the steady-state consonant is manipulated. There are three different sequences, depending on the coarticulation degree between F2 frequency in the lateral consonant and in the following vowel.

Test 3: Both F2 frequencies in the steady-state of [l] and in transitions are manipulated in three different sequences.

Fifteen items were generated to study the three possible sequences obtained for [l] closed to the five Spanish vowels. These recorded stimuli were grouped in thirty pairs in random order. The same method was applied in the three tests. A panel of 40 listeners (native Spanish speakers, about 20 years old, students at the Alicante University) were asked to choose the best item in each pair. The whole serie had 90 pairs. They could listen to each one twice.

3.RESULTS

The results of the perceptual tests are shown in Figure 1 below:

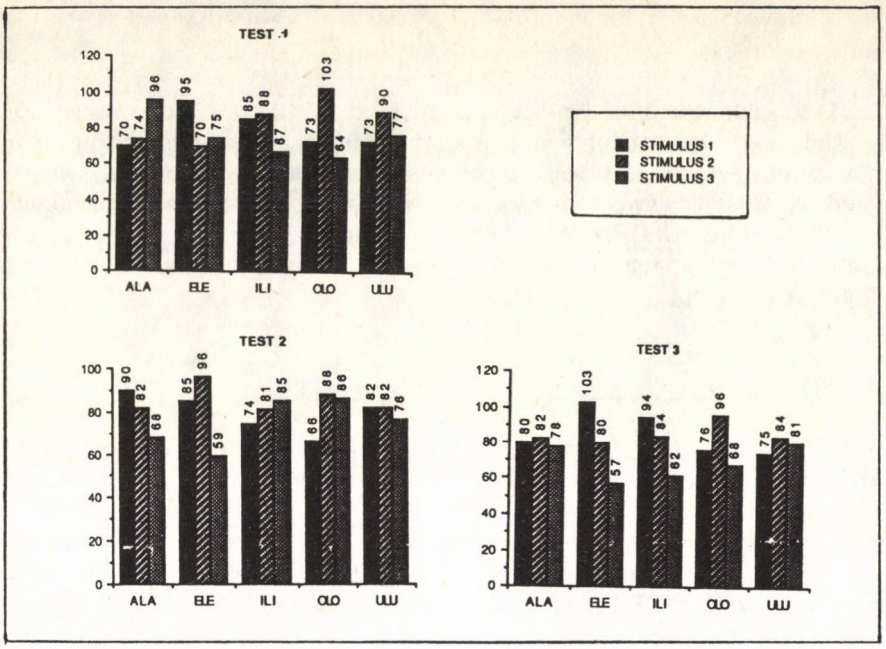


Figure 1.

Some remarks can be made about this data:

a) Stimulus 2 is generally the preferred one in the three tests. In test 1 stimulus 2 was considered the best one in [ili], [olo], [ulu]; stimulus 3 was the most preferred in [ala], and stimulus 1 in [ele]. In test 2 stimulus 2 was also the best one in three cases ([ele], [olo], [ulu]); stimulus 3 was the best one in one case [ili], and stimulus 1 in [ala]. Finally, in test 3, stimulus 2 was the best in [ala], [olo], [ulu], and stimulus 2 in the other cases ([ele], [ili]).

b) Further, we can see that stimulus 2 is also the best one both in the first and the second test, if we consider the results for five vowels together (425 and 429 answers, respectively); in test 3, however, stimuli 1 and 2 were more or less at the same level (428 answers for stimulus 1 and 426 for stimulus 2).

c) The preference for stimulus 2 is clear in any of the three tests whereas the vowel after [l] is back ([o],[u]). When the vowel is [i] or [e] the number of choices for stimulus 1 is greater than in back vowels (stimuli with [e] are the preferred ones). When the vowel is [a], however, there is no preference for any stimulus.

d) The differences between stimuli are clearer in test 1. These differences are the same in test 3 . This seems to indicate the importance of transitions as coarticulatory perceptual cues.

4.CONCLUSION

The results of these tests suggest that there is a tendency to prefere items in which there is a certain degree of coarticulation generated with intermediate values between [l] and the vowel. This is specially the case in [o] and [u]. Nevertheless, for [i] and [e] the degree fo coarticulation seem to be less important. An excessive coarticulation is related to a low identification score. It also seems that changes in F2 transitions are more important for the perception of lateral consonants than changes of the second formant in the steady-state of the consonant.

As we were working on the role of coarticulatory phenomena in [l] perception, the tendencies observed in our results also suggest that variations in the degree of coarticulation do not seem to be a primary factor in the correct identification of the consonant. However, this element is important from the point of view of naturalness.

REFERENCES

[1] AINSWORTH, W.A (1968) "First formant transitions and the perception of synthetic semi-vowels" *JASA*, 44, 3, pp. 689-694.

[2] CHAFCOULOFF, M. (1980) "Les caracteristiques acoustiques de [j, y,w, l, r] en français" *Travaux de l'Institut de Phonetique d'Aix*, Vol. 7, 1-52.

[3] CHAFCOULOFF, M. (1983) "A propos des indices de distinction [l-r] en français" *Speech Communication*, 2, pp. 137-139.

[4] LISKER, L. (1957) "Minimal cues for separating [w, r, l, j] in intervocalic position" *Word*, 13, 2, pp. 257-267.

[5] O' CONNOR *et al.* (1957) "Acoustic cues for the perception of initial [w, j, r, l]" *Word*, 13, pp. 24-43.

LEFT AND RIGHT CONTEXT EFFECTS IN SPEECH PROCESSING

Uli H. FRAUENFELDER & Anna M. SOUVERIJN
Max-Planck-Institut für Psycholinguistik
Nijmegen, NL.

Introduction

Spoken language understanding requires the perceptual integration of temporally distributed information at several levels. Listeners must extract and integrate acoustic cues that are scattered across relatively long stretches of speech to construct segmental percepts. Similarly, they must analyze and integrate these overlapping segments to recognize words. Any account of these processes must explain how and when the integration of different information is accomplished at each level and between the different levels.

Considerable experimental evidence suggests that these integrative processes take place immediately and continuously. At the phoneme level, Repp (6) has shown that listeners can extract phonetic information continuously and can make phonemic decisions before all the necessary cues for the segment have become available. Similarly, at the lexical level, Warren and Marslen-Wilson (8,9) employed the gating procedure to demonstrate that partial sensory information about incoming segments is projected onto the lexical level without apparent delay. This assumed close relationship between the arrival of the speech signal and its analysis leaves little room for either **right context** effects in which later arriving information influences earlier perceptual decisions or **left context effects** in which earlier information affects perceptual decisions about later arriving segments.

In this paper we look for such context effects to confirm or infirm the immediate and on-line character of the speech processing that is reflected by the phoneme detection task (1). To do so, we manipulated both the left and the right context of the phoneme targets that the subjects were to detect. In particular, we varied the properties of the (critical) phonemes that served either as the left or the right context for the target phoneme. Since critical phonemes were separated in both cases by a vowel, they presumably carried few, if any, acoustic cues directly relevant to the perception of the target phoneme. Nonetheless, these context phonemes were potentially relevant for the response as we will see.

Right Context effects.

Our objective in this first experiment was to determine whether the detection of a target phoneme could still be influenced by a following phoneme. Subjects were asked to detect phoneme targets (C1) in monosyllabic words and nonwords with C1V1C2 structure. We varied the phonological distance distinguishing context (C2) and target phonemes. In the Identical condition, both phonemes were the same (/p/ as in "pap"), whereas in the Different condition, they differed by several distinctive features (/p/ versus /l/ as in "pal"). Evidence for right context effects were expected to take the form of reaction time (RT) differences between the Identical and the Different conditions. Indeed, diverse models (4) predict that the redundant critical phoneme (/p/) should accelerate the detection of the identical target phoneme. The physical distance between the target and critical phoneme was also manipulated by varying the phonemic length of the vowel, V1 (e.g., /a/ versus /aa/). In this way, we hoped to determine the size of the temporal window during which such right context effects might operate. It was assumed that the later the critical phoneme arrived, the smaller its impact would be, leading to a smaller facilitatory effect of the critical phoneme for longer vowels. Finally, the lexical status of the target-bearing item was varied to establish whether the effects were lexically mediated. If right effects

are obtained for both words and nonwords, they cannot simply be attributed to a greater lexical contribution to the detection of phonemes in words with the redundant phonemes.

Subjects – Thirty-five native speakers of Dutch participated in this study.

Materials – Eighteen pairs of Dutch monosyllabic words and nonwords with C1V1C2 structure were used in this experiment. The target phonemes were /p/ and /k/. Examples of stimuli for the target /k/ are shown in Table 1.

Table 1

Lexical Status	Vowel Length	Phonological Distance	
		Identical	Different
Word	Short	kok	kol
Word	Long	kook	kool
Nonword	Short	kuk	kum
Nonword	Long	keuk	keung

The target-bearing items were embedded in six counterbalanced lists made up of other words and nonwords of different length. Each list contained 60 items: 4 test items, 26 target-bearing fillers and 30 items non-target bearing items. For each list, subjects were asked to detect one of the two previously specified targets that could occur anywhere within the target-bearing item (c.f. 3 for more details about this procedure) as quickly as possible.

Results

Mean reaction times (measured from the burst of targets) were computed for each subject and each experimental item. All responses less than 100 ms. or greater than 1000 ms. were not included in the computation of the means. Table 2 shows the RTs (in msec.) and percent errors for Identical and Different conditions broken down as a function of lexical status (words versus nonwords) and vowel length (short versus long).

Table 2

		Phonological Distance					
		Identical		Different		Difference	
		RT	error	RT	error	RT	error
Word	Short	410	0.4	440	5.7	30	5.3
Word	Long	457	2.2	456	7.9	-1	5.7
Nonword	Short	417	0.0	443	2.6	26	2.6
Nonword	Long	411	0.4	462	5.7	51	5.3

An analysis of variance showed that the main effects of phonological distance and vowel length were both significant by subject ($F_1(1,37) = 28.6$; $p < .001$; $F_1(1,37) = 112.9$; $p < .001$) and the former was also marginally significant by item ($F_2(1,3) = 7.17$; $p < .07$). Furthermore, there was a weak interaction between the effect of the phonological distance and the lexical status of the target-bearing item ($p < .02$). Finally, separate planned comparisons showed that effects of phonological distance were significant for both words and nonwords.

The results of this experiment show that the detection of item-initial targets is influenced by their right context. Both words and nonwords showed faster detection latencies

when the target was followed by an identical phoneme. The manipulation of the physical distance separating target and critical phoneme produced less clear results. Although words showed the predicted greater phonological distance effect for the short vowel condition, unexpectedly, the reverse was true for nonwords. The results for the nonwords show that context phonemes, arriving even several hundred msec after the target (i.e., the vowel duration varied between 150 and 250 msec), can still influence the detection response.

Left Context effects.

We were interested not only in right but also in left context effects where the critical context phoneme precedes rather than follows the target. In the Similar condition, the critical phoneme and target phonemes are similar (e.g., "teipu" where the target is /p/), whereas in the Dissimilar condition, they differ by several distinctive features ("leipu"). Previous phoneme monitoring research (2,5,7) has shown longer detection latencies for the Similar condition than for the Dissimilar condition when the critical and target phonemes were in different words produced in sentences. In this experiment, both critical and target phonemes are in the same item separated by only a vowel.

Subjects – Thirty native speakers of Dutch participated in this study.

Materials – Eighteen pairs of pronounceable nonwords, all with C1V1C2V2(C3) structure, made up the test materials. In the first member of the pair (the Similar condition), the target phoneme (C2) differed from the preceding critical context phoneme (C1) by a single distinctive feature, whereas in the second (Dissimilar condition) the target and left context phoneme differed by at least 3 distinctive features. The target phoneme was either /p/, /t/, and /k/.

The target-bearing items were embedded in nine counterbalanced lists made up of other words and nonwords of different length. Each list contained 27 items: 4 test items, 4 target-bearing fillers and 20 non-target bearing items. Subjects were asked to detect targets that could appear anywhere in the target-bearing item.

Results

Mean reaction times (measured from the burst of the targets) were computed for each subject and each experimental item. All responses less than 100 ms. or greater than 1000 ms. were not included in the computation of the means. Table 3 presents examples of the stimuli used (with the /p/ target) and reaction times and errors for Similar and Dissimilar conditions.

Table 3

Phonological Distance

	Similar	Dissimilar	Difference
Stimuli	teipu	leipu	
RTs	389	364	25
Errors	3.1%	1.1%	2.0%

An analysis of variance revealed that the effect of phonological distance was significant by subject ($F_1(1,29) = 15.8$; $p < .001$) and by item ($F_2(1,16) = 3.46$, $p < .09$). The difference in the error rate for Similar and Dissimilar conditions was also significant.

These results have shown clear effects of left context. Detection latencies were slower when the initial phoneme was similar to the target phoneme. Diverse explanations have been advanced for left context effects, including false alarming (5), greater processing cost for determining that the similar critical phoneme is not the target (4), and greater inhibition of the target phoneme by a similar phoneme (7). Unfortunately, our results do not provide a firm basis for deciding between these different interpretations.

Conclusion

The results of the two experiments presented here have demonstrated that phoneme detection latencies are sensitive to both left and right context. The results showing the right context effects suggest that the speech processing required in the detection of phonemes is not as immediate as is generally thought, but rather is highly sensitive to the following context. In addition to their obvious consequences for the contruction of experimental stimuli, these results raise some important questions concerning the locus of the context effects. We need to establish whether these effects are perceptual or whether they are due to the specific response elicited in the task. Models like TRACE (4) can provide a unitary perceptual mechanism based upon inhibition and facilitation between adjacent phonemes to account for both effects. Nonetheless, these results do not allow us to entirely reject alternative explanations - like one appealing to statistical facilitation from the right context phoneme. Futher research is clearly required to tease apart these different explanations.

References

1. Cutler, A. & Norris, D.: Monitoring sentence comprehension. In W.E. Cooper & E.C.T. Walker (Eds.). *Sentence processing: Psycholinguistic studies presented to Merrill Garrett. Hillsdale* N.Y.: LEA, 1979.
2. Dell, G.S. & Newman J.E.: Detecting phonemes in fluent speech. *Journal of Verbal Learning and Verbal Behavior*, 19, 1980, 608-623.
3. Frauenfelder, U.H. & Segui, J.: Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory and Cognition*, 17,2, 1989.
4. McClelland, J. L., & Elman, J. L. . The TRACE model of speech perception. *Cognitive Psychology*, 18, 1986, 1-86.
5. Newman, J.E. & Dell, G.S.: The phonological nature of phoneme monitoring: A critique of some ambiguity studies. *Journal of Verbal Learning and Verbal Behavior*, 17, 1978, 359-374.
6. Repp, B. H.: Accessing phonetic information during perceptual integration of temporally distributed cues. *Journal of Phonetics*, 8, 1980, 185-194.
7. Stemberger, J.P., Elman, J.L. & Haden, P.: Interference between phonemes during phoneme monitoring: Evidence for an interactive activation model of speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 11(4), 1985, 475-489.
8. Warren, P. & Marslen-Wilson, W.D.: Continuous uptake of acoustic cues in spoken word recognition. *Perception & Psychophysics*, 41, 1987, 262-275.
9. Warren, P. & Marslen-Wilson, W.D.: Cues to lexical choice: Discriminating place and voice. *Perception & Psychophysics*, 43, 1988, 21-30.

ZUR PERZEPTION PHONETISCH ABWEICHENDER SPRACHE VON AUSLÄNDERN

Ursula HIRSCHFELD
Herder-Institut der Karl-Marx-Universität
Leipzig, Deutsche Demokratische Republik

Problematik

Untersuchungen im Bereich der Technik mit synthetischer oder synthetisch veränderter Sprache bzw. im Bereich der Logopädie mit Sprach-, Sprech- und Hörstörungen haben interessante Ergebnisse für die Perzeption phonetisch abweichender Sprache gebracht (Bedeutung des situativen, semantischen und lautlichen Kontextes und der Redundanz (6), Leitfunktion des prosodischen bzw. suprasegmentalen Bereichs (5)). Sie dürfen aber nicht, wie es häufig geschehen ist, unkritisch als übertragbar oder als Ersatz für Untersuchungen zur Perzeption der phonetisch abweichenden Sprache von Ausländern betrachtet werden. Gerade die Kontexthilfe und die Redundanz werden oft als Argument benutzt, phonetische Fehlleistungen in der Zielsprache zu bagatellisieren. Dabei wird übersehen, daß - im Gegensatz zu den sprachlichen Veränderungen in Technik und Logopädie, die meist systematisch, erkennbar und somit voraussagbar sind - Fremdsprachenlernende im gewissen Sinne unsystematische, nicht voraussagbare und komplexe, d. h. alle Sprachebenen betreffende, sowie verschiedene (emotionale, die Konzentration beeinträchtigende usw.) Nebenwirkungen provozierende Abweichungen hervorbringen. Für den naiven Hörer kommt es nicht nur zu fehlenden, sondern auch zu vertauschten oder zusätzlichen Informationen; der gestörte Kontext (falsche Lexik, Syntax- und Grammatikfehler) bietet oft keine Hilfe, er belastet die Verständlichkeit zusätzlich.

Es ist ohne Zweifel notwendig, spezielle Untersuchungen zur Perzeption solcherart veränderter Sprache vorzunehmen. Erste Ansätze verweisen auch hier auf die besondere Bedeutung der Prosodie (1, 3). Die Perzeptionsergebnisse sind zudem vom Hörer mit seiner Motivation und seinen individuellen Fähigkeiten stark abhängig (2). Um den insgesamt unbefriedigenden Untersuchungsstand weiter zu verbessern und um schließlich eine Konkretisierung von Ziel und Inhalt des Phonetikunterrichts im Fach Deutsch als Fremdsprache vornehmen zu können, wurde die phonetische Verständlichkeit von etwa 60 ausländischen Studenten bei etwa 800 deutschen Hörern untersucht. Für die Perzeption der in Abhängigkeit von der Muttersprache spezifisch abweichenden deutschen Aussprache ergibt sich ein kompliziertes, vielschichtiges Bedingungsgefüge. Zu konstatieren ist ein erhöhter Aufwand bei der Verarbeitung dieser im Vergleich zur Norm und zum Gewohnten erheblich veränderten Sprache. Er zeigt sich deutlich in Informationsverlusten und einer verminderten Behaltensleistung beim Hörer, in Konzentrationsschwächen und einer vom Inhalt auf die Form gelenkten Aufmerksamkeit. Er zeigt sich auch in der verbalen Einschät-

zung der Aussprache verschiedener Sprecher durch die Hörer, die mittlere und schlechte phonetische Leistungen u. a. zunehmend als mangelhaft, unangenehm, unzulänglich und ungewohnt bewerteten (4). Hauptziel dieser Untersuchungen war es, die Wirkung von Abweichungen in der segmentalen und suprasegmentalen Struktur (distinktive Merkmale der Vokale und Konsonanten, Stellung und Zusammensetzung der Akzentsilbe, rhythmische Struktur) zu bestimmen. Neben mehreren weiteren, hier nicht zu referierenden Versuchen zur Satz-, Wortgruppen- und Wortverständlichkeit, in denen die Hörer unter möglichen, vorgegebenen Hörergebnissen das Verstandene zu markieren hatten, wurde im nachfolgend beschriebenen Experiment ohne Vorlage von Minimalpaaren gearbeitet.

Methode

Folgende Versuchspersonen waren beteiligt: 1. als Sprecher ein nikaraguanischer Student nach achtmonatigem Sprachunterricht, mit durchschnittlichen phonetischen Leistungen (3,0 auf der Zensurenskala von 1 bis 5) und für spanischsprechende Deutschlernende typischen Ausspracheabweichungen; 2. als Hörer 40 Muttersprachler unterschiedlichen Alters (sechs Gruppen mit je fünf Hörern zwischen 8 und 65 Jahren) und unterschiedlicher Berufe, sowie je fünf Phonetik- und Sprachlehrer.

Das Testmaterial bestand aus 52 Wörtern, 52 Familiennamen und 22 Logatomen (vom Typ [m-V-nə]), die alle für das Deutsche wesentlichen Distinktionen enthielten und in ihrer lautstatistischen Zusammensetzung bis auf geringe Abweichungen den für das Deutsche ermittelten Werten (7) entsprachen. Alle Testbeispiele waren zweisilbig, der Akzent lag in der Regel auf der ersten Silbe. 90 % der Wörter gehörten zu einem Mindestwortschatz von 3000 Wörtern, waren also den Versuchspersonen aller Altersgruppen geläufig.

Die Tonaufnahmen (Studioqualität) wurden von mehreren Phonetikern transkribiert, Abweichungen wurden nach einem vereinbarten System erfaßt und kategorisiert. Die deutschen Versuchspersonen hörten die Aufnahmen einzeln und beliebig oft; sie sollten notieren, was sie verstanden hatten.

Ergebnisse

1. Die AKZENTSTELLE blieb in den Hörurteilen (HU) fast generell erhalten, auch wenn starke Abweichungen im segmentalen Bereich auftraten. Bei einander ausschließender Kombination von Akzent und lautlicher Realisation - wie z. B. in [ge'vek], wo der Akzent eher dem Wort "Gebäck", die segmentale Ebene eher dem Wort "Gehweg" entspricht - erweist sich der Akzent als stärker, als für die Perzeption wesentlich. In allen 2080 HU bei den Namen und in 2075 von 2080 HU bei den Wörtern blieb die Stellung des Wortakzents erhalten.

2. Die SILBENZAHL erwies sich ebenfalls als eine wichtige und stabile Größe. Bei allen Namen blieb sie entsprechend

der Realisation bestehen; bei den Wörtern zeigte sich in 6 der 2080 HU eine veränderte, und zwar verminderte Silbenzahl.

3. Die STRUKTUR DER AKZENTSILBE, die Aufeinanderfolge von Vokal und Konsonant(en) - ungeachtet dessen, ob der realisierte Laut korrekt wahrgenommen wurde-, veränderte sich in 8,8% der Wörter, dabei kam es sowohl zum Verlust (KKV, KVK → KV) als auch zum Zusatz (KV → KKV, KVK) von Konsonanten. Oft bestimmte die Realisation des Vokals, teilweise auch in Verbindung mit dem initialen Konsonanten, das Perzeptionsergebnis - immer im Zwang der Hörer, ein sinnvolles Wort zu erkennen. Bei den Namen gab es deshalb wesentlich weniger Veränderungen in der Akzentsilbenstruktur.

4. Die STRUKTUR DER NACHAKZENTSILBE variierte demgegenüber stärker; sie mußte entsprechend der wahrgenommenen Akzentsilbe zurechtgehört werden, damit ein sinnvolles Wort, ein geläufiger Name entstand. In einigen Fällen waren mehrere verschiedene, ähnlichklingende Endsilben möglich, die von den Hörern auch genannt wurden. Das mag darauf hinweisen, daß die Endung in ihrer konkreten Form eine untergeordnete Rolle spielt und durch den Kontext determiniert wird. Den 52 gehörten Varianten entsprechen bei den Wörtern 213 und bei den Namen 183 (z. T. mehrmals) angegebene.

5. Veränderungen bei den DISTINKTIVEn MERKMALEn DER VOKALE wirkten sich bei den Wörtern vor allem dann aus, wenn die jeweilige Opposition ein Minimalpaar ergeben könnte. Bei den Namen und Logatomen spielte der lautliche Kontext nicht diese Rolle, hier wurde die direkte Wirkung der Abweichungen besser sichtbar. Während bei den Wörtern nur drei Fälle auftraten, in denen veränderte distinktive Merkmale (und zwar Qualität/Quantität) eine Entscheidung der Hörer für eins von zwei möglichen Wörtern erforderte - in 102 von 120 HU dominierte die Quantität -, war dieser Zusammenhang bei den Namen generell gegeben, man denke an Varianten wie Möhler, Möller, Mohler, Mehler usw. Für die Namen und Logatome ergab sich folgende Substitutionsmatrix (Ausschnitt):

REALISATION	HÖRURTEILE (absolut)							Summe
	e:	ɛ	ɛ:	i:	ɪ	ø:	œ	
e:	65	5	9	--	--	1	--	80
e·	16	34	23	--	5	2	--	80
e	8	77	2	1	21	--	11	120
ɛ:	23	--	16	--	--	1	--	40
ɛ	--	--	40	--	--	--	--	40

Diese Matrix wurde nach den einzelnen, im HU erhalten gebliebenen distinktiven Merkmalen aufgeschlüsselt; für die oben dargestellte Realisation [e] ergibt sich z.B. folgendes Bild:

MERKMAL	REALISATION	HÖRURTEILE
Quantität	kurz	109 kurz - 11 lang
Qualität	geschlossen	9 geschlossen - 111 offen
Rundung	ungerundet	109 ungerundet - 11 gerundet
Hebungsgrad	mittel	98 mittel - 22 hoch
Richtung	vorn	120 vorn

Die Analyse aller realisierten Vokale in den Namen und Loga-

tomen brachte folgende Ergebnisse:
- die Quantität dominierte eindeutig mit 90,1 % bei den Namen und mit 90,3 % bei den Logatomen; die Quantität blieb in 42,9 % der Namen bzw. 34,5 % der Logatome erhalten;
- die Lippenrundung blieb in 97,9 %,
- der Hebungsgrad der Zunge in 95,4 %,
- die Hebungsrichtung in 98,6 % aller HU erhalten.

6. Die DISTI"KTIVEn MERKMALE DER KONSONANTEN wurden in gleicher Weise überprüft. Substitutionsmatrix und Analyse ergaben:
- die Artikulationsart blieb in 95,6 %,
- die Spannung (fortis-lenis) in 93,1 %,
- die Artikulationsstelle in 88,2 % aller HU erhalten.

Zusammenfassung

Die genannten Ergebnisse werden in ihrer Tendenz von den anderen zur Untersuchung gehörenden Experimenten bestärkt und lassen (vorsichtige) Schlußfolgerungen zur Perzeption phonetisch abweichender Sprache Deutschlernender zu. Sie zeigen, daß dem Akzent und der Struktur der Akzentsilbe sowie der Realisation des Akzentvokals besondere Bedeutung zukommt. Das Perzeptionsergebnis wird durch Abweichungen in der Vokalquantität stärker beeinträchtigt als bei qualitativen Veränderungen. Bei den Konsonanten wirken sich vor allem Ungenauigkeiten in Artikulationsart und Spannung negativ aus. Die Versuche zeigen weiter, daß der Kontext vielfach nicht die ihm i. allg. zugesprochenen Hilfsleistungen bringt; er wird vielmehr der konkreten Realisation angepaßt.

Bemerkenswert ist auch die beobachtete und nachgewiesene indirekte Wirkung unkorrekter Äußerungen - Konzentrations-, Aufmerksamkeitsverluste, negative Emotionen beim Hörer.

Weitere Untersuchungen sind erforderlich, um diese Zusammenhänge zu konkretisieren und zu vervollständigen.

Literatur

1. BANNERT, R.: Intelligibility of Foreign Accent. Abstracts of the 10th ICPS. 1984, 600.
2. DAVIES, E. E.: Error Evaluation: the Importance of Viewpoint. ELT-Journal. 37/4/1983, 304--11.
3. EISENSTEIN, M.: Negative Reactions to Non-native Speech. Studies in Second Language Acquisition 5/2/1983, 160-76.
4. HIRSCHFELD, U.: Faktoren der phonetischen (Wort)Verständlichkeit. Proceedings 11th ICPS vol. 4. 1987, 221--4.
5. HUGGINS, A.: On the Perception of Temporal Phenomena in Speech. J.A.S.A. 51/4/1972, 1279--90.
6. HUNNICUT, S.: Intelligibility vs. Redundancy - Conditions of Dependency. Lang. and Speech. 28/1/1985, 47--56.
7. MEINHOLD, G.--STOCK, E.: Phonologie der deutschen Gegenwartssprache. Leipzig, 1980.

PREDICTING PERCEPTUAL CONFUSIONS IN SWEDISH VOICED STOPS

Diana KRULL
Institute of Linguistics
University of Stockholm
S-106 91 Stockholm, Sweden

Introduction

In speech recognition algorithms and certain theories of speech perception the interpretation of the signal is based on "distance scores" for comparisons of the signal with stored references. In these theories, perception is seen as a product of stimulus and experience. (1),(2),(3). The aim of the present paper is to evaluate such distance measures by investigating listeners' confusions of the Swedish voiced stops in variable vowel context. To what extent can perceptual identifications be accounted for in terms of the acoustic properties of the stimuli?

Perceptual data

To elicit perceptual confusions, stimuli were constructed using a $V_1C:V_2$ frame, with the Swedish voiced stops */b,d,ɖ,g/* in intervocalic position; the vowel context consisted of all 25 possible combinations of phonologically short variants of */i,e,a,o,u/*. The resulting 100 stimuli were read in random order by a male phonetician using the Swedish grave accent in which both syllables have about equal prominence. Additional stimuli were prepared on the basis of the original recordings. Firstly, the stimuli were cut into two halves, V_1C and CV_2. Secondly, stimulus fragments of ca 26 ms, beginning just before consonant release, were cut out. These short stimuli will be referred to as "burst" although they may also contain the beginning of the following vowel.

The four sets of stimuli were transferred to different tapes. Each stimulus occurred three times, with the exception of the original (longest) ones which occurred once. Twenty normal–hearing native speakers of standard Central Swedish listened to the stimuli, their task being to identify the consonant. The results of the listening test showed that that the CV_2 stimuli were identified almost as well as the original stimuli. The few confusions that occurred were almost exclusively between the dental and the retroflex consonants. V_1C stimuli were more difficult, but the greatest number of confusions occurred in the "burst" stimuli. The prediction models in this paper deal only with these stimuli.

The confusions, shown in Fig.1, formed a regular pattern depending on the vowel context. In particular, the asymmetries in the confusion matrix varied systematically with the front–back dimension of the following vowel. For example, front V_2 favored dental and retroflex answers, and back V_2 labial and velar. Of the dental–retroflex pair, the retroflex received a higher score with back V_2. The asymmetries were regular enough to be predictable given F_2 at CV boundary and in the middle of the following vowel (cf.(4) p.109).

Acoustic data

The acoustic data used to predict the above perceptual confusions were based on differences in formant frequencies, spectrum levels, and the duration of the noise section following consonant release. F_2, F_3 and F_4 were measured at the moment of consonant release on wide band (300 Hz) spectrograms. The results transposed from Hertz to Bark using a method presented by Traunmüller (5). F_2, F_3 and F_4 were regarded as dimensions in a three–dimensional space where each stimulus was represented as a point. Distances between points were calculated according to the Euclidean metric using the equation

$$\mathrm{Dform}_{i,j} = \sqrt{(\Delta F_2)^2 + (\Delta F_3)^2 + (\Delta F_4)^2} \qquad Eq.1$$

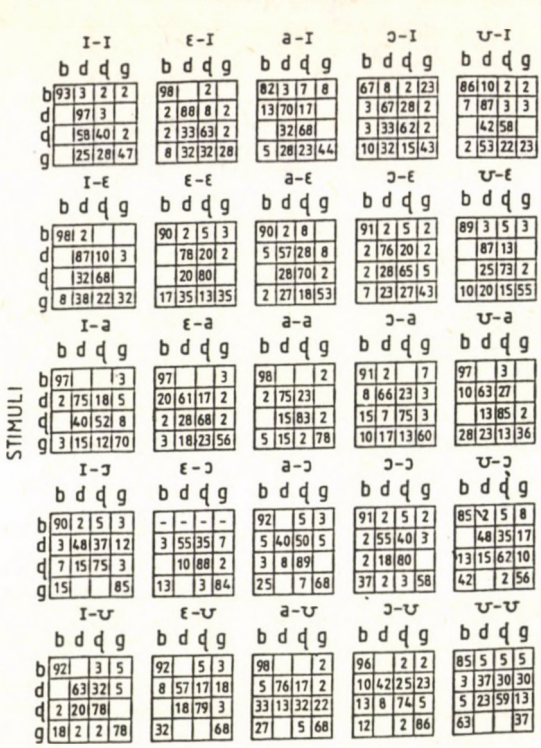


Figure 1 *Confusion matrices for "burst" stimuli in 25 vowel contexts.*

where Dform is the formant–based distance at the CV boundary and Fn is the critical band rate in Bark of the n^{th} formant. The duration of the noise section after consonant release (burst length) was measured on oscillograms, from the moment of consonant release to the first noise free pulse of the vowel. For details see (4). Differences between the burst lengths were calculated according to the equation

$$Dbi_{i,j} = | B_i - B_j | \qquad Eq.2$$

where Db is the difference between the burst lengths B of the stimuli i and j. Distances between spectra were calculated using 14 non–overlapping bandpass filters with the width of 1/4 of an octave and center frequencies from 446 Hz to 4243 Hz. SPL was measured at two points in time: t_1 integrated over the 10 ms following consonant release, and t_2 10 ms later. Distances between spectra at t_1, referred to as static distances, were calculated using the equation

$$Dstat_{i,j} = \sqrt{\sum_{n=1}^{14} |L_{i,n} - L_{j,n}|^2} \qquad Eq.3$$

where $L_{i,n}$ and $L_{j,n}$ are the levels in dB of the stimuli i and j at the n^{th} filter band. To assess differences in dynamic change between t_1 and t_2, dynamic distances, the equation used was

$$Ddyn_{i,j} = \sqrt{\sum_{n=1}^{4} |C_{i,n} - C_{j,n}|^2} \qquad Eq.4$$

where $C_{i,n}$ denotes the level change in stimulus i, band n. Finally, the static and dynamic distances were combined

$$D_{i,j} = \sqrt{(Dstat_{i,j})^2 - (Ddyn_{i,j})^2} \qquad Eq.5$$

where Dstat is the static distance (Eq.3) and Ddyn, the dynamic distance (Eq.4) and i,j different stimuli.

Prediction of the perceptual responses

Plotting confusions vs. the different kinds of acoustic distances described above produced very noisy graphs although there was always a decrease of confusions with increasing distance. Two possible reasons for this dispersion may have been that the listeners' responses were asymmetrical or that the wrong reference values (prototypes) were used when calculating the distances. The responses were therefore symmetrized by using mean values of the responses of a pair of stimuli, giving slightly better predictions of the confusions. (The

asymmetries were restored later.) Next, the prototypes were investigated. The acoustic distances had been calculated between stimuli with the same vowel context, e.g. between */ibu/* and */igu/*, etc. An extra listening test with the "burst" stimuli where the listeners were asked to identify the vowel showed that V_2 was often difficult to identify. Especially, */o/* and */u/* after a dental or retroflex consonant were identified as front vowels. After experimenting with different prototypes it appeared that /VCa/ gave the best predictions except for */b/* and */g/* before back vowels. In the latter case, prototypes with the original vowel gave the best results.

There was little difference between the formant based and the spectrum based distances regarding prototypes. In back vowel context, however, the formant model did not always predict the confusions between */b/* and */g/* successfully: there were often few confusions when the calculated distance was small. To explain this effect, the difference in burst length was added to the formant based distances using the equation

$$Dm_{i,j} = \sqrt[p]{(w_1 * Dform_{i,j})^p * (w_2 * Db_{i,j})^p} \qquad Eq.6$$

where Dm is the modified distance, Dform is the formant based distance and Db is the difference in burst length, i and j are different stimuli, w_1 and w_2 are weighting factors and p is a constant. In this case, $w_1 = 1$, $w_2 = .1$ and $p = 2$ gave the best correlation between the calculated distance and observed confusions. The modified distances improved the predictions, especially for stimuli with back V_2. Table I shows some of the correlation coefficients for the different models.

Table I. *Correlation coefficients for percent confusions as a function of different calculated distances: I. Formant–based distance with reference V_2 = mean of stimulus V_2 (column a), or V_2 = /a/ except for /b/ and /g/ followed by /o/ or /u/ where reference V_2 = mean of stimulus V_2 (column b); II. Spectrum based distances according to Eq.1, reference V_2 as in I (column a and b). The numbers in parentheses give the coefficients of Spearman rank–order correlation.*

	I.		II.	
	(a)	(b)	(a)	(b)
Stim. V_2				
/i,e/	*–.41*	*–.71 (–.77)*	*–.40*	*–.63 (–.60)*
/a/	*–.63*	*–.63 (–.65)*	*–.63*	*–.63 (–.64)*
/o,u/	*–.74*	*–.63 (–.59)*	*–.27*	*–.57 (–.53)*

Of the spectrum–based models the dynamic one (Eq.4) was more efficient than the static, the combined model (shown in Table II) falling in between the two. The formant–based model gave the highest correlations with listeners' confusions, and it could be further improved. As shown by Fig.2, there is a linear decrease in the number of confusions with calculated distance only up to a certain point – further increase in the distance does not affect the number of confusions. New regression analyses were performed only to the distance of 2.3 Bark for V_2 = */i,e,a/* and 3.2 Bark for V_2 = */o,u/*, for the larger distances the median value of the confusions was used as a prediction.

The predictions described applied to the symmetrized percent confusions. However, for each consonant pair Ci,Cj the quotient between Ci–answers and Cj–answers could be predicted given F2 of V2 at the CV boundary and in the middle of the vowel ((4) p.108). When, finally, the non–symmetrized percent confusions were plotted against the new calculated distances, the correlation coefficient was $r = .85$.

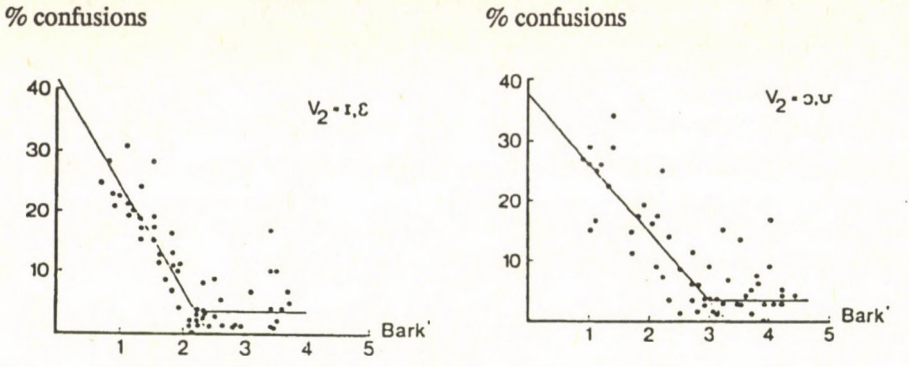


Figure 2 *Percent confusions as a function of the combined distance measure based on formant frequencies and burst length See text for details.*

Conclusions

Listeners' perceptual confusions can to a large extent be explained by acoustic properties with the addition of minimal assumptions about auditory transformation. For this purpose, the most reliable acoustic properties turned out to be the formant frequencies at the CV boundary, combined with the noise section after consonant release. The best correlation between predicted and observed percent confusions was r = .85. It may not be possible to arrive at a much higher correlation between predicted and observed percent confusions because the number of confusions contained considerable inter–subject variation.

References

1. CARLSON, R. and GRANSTRÖM, B.: Model predictions of vowel dissimilarity. *STL-QPSR 3-4/1979*, Royal Institute of Technology, Stockholm, 1979.

2. BLADON, R.A.W. and LINDBLOM, B.: Modeling the judgement of vowel quality differences. *J.Acoust. Soc. Am.*69(5), 1981, 1414–1422.

3. POLS, L.C.W.: Variation and interaction in speech. *Report nr. 74*. Institute of Phonetic Sciences, University of Amsterdam, 1983.

4. KRULL, D.: *Acoustic properties as predictors of perceptual resposes. A study of Swedish voiced stops*. Phonetic Experimental Research at the Institute of Linguistics University of Stockholm (PERILUS), Stockholm, 1988.

5. TRAUNMÜLLER, H.: Analytical expressions for the tonotopical sensory scale. (Part of a doctoral dissertation). Institute of Linguistics, University of Stockholm, 1983.

INTELLIGIBILITY AND NATURALNESS OF SYNTHETIC CV WITH VARYING DEGREES OF COARTICULATION

Joaquim LLISTERRI, Gemma MARTINEZ-DAUDEN and Natividad FERNANDEZ-GUTIERREZ
Laboratori de Fonètica, Facultat de Lletres, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

Almost five years ago, J. Allen (1985: 1546-7) stated that "there are no contemporary speech synthesis systems that begin to approach the level of surface phonetic variability observed in natural speech". Despite of the great improvements that have been made in text-to-speech conversion since that time, Allen's assertion still holds true. However, there is no doubt that it is possible to produce high-quality synthetic speech with some of the available text-to-speech systems (e.g. MITalk, Prose 2000, DECtalk or Infovox); their performance in different tests yields a high degree of intelligibility and it is expected that "the speech generated by these high quality systems may soon approach the almost perfect levels of intelligibility observed for natural speech" (Pisoni, Nusbaum and Greene, 1985).

The research into the intelligibility of text-to-speech products has led to the question of the improvements that can be made in such systems. Among them, many factors related to what is usually called phonetic implementation rules seem to be crucial in determining the segmental quality of synthesized speech. It has been shown in a system like MITalk that low error rates in a phoneme recognition task may be obtained when special attention is paid to the modelling of CV transitions in the rules of the phonetic component (Allen, Hunnicutt and Klatt, 1987). These rules are derived from the acoustic analysis of natural utterances, following a procedure that is known as "analysis by synthesis"; phonetic information is usually obtained from a careful study of one speaker recorded in laboratory conditions reading CV, CVC or CVCV nonsense words.

One of the major problems we face when aiming at deriving synthesis rules from the acoustic analysis of natural

utterances is the modelling of stop-vowel transitions: stop consonants tend to coarticulate with the following vowel and some strategies have to be applied to obtain a natural transition from consonant release to vowel onset. In systems like MITalk this is done with a " locus equation " that predicts F1, F2 and F3 values at voicing onset from a knowledge of the vowel target.

There is no need to say that the concept of "locus" is well known in speech synthesis studies since the classical contribution by Delattre, Liberman and Cooper (1955) in which the importance of F2 locus was clearly emphasized. However, as Delattre (1969) himself pointed out some years later, in the acoustic analysis of natural speech the " locus " must be distinguished from the " convergence point " of formant transitions, the later being dependent on the following vowel. It seems then that there is a certain amount of disagreement between the concepts used in the analysis of natural speech and the elegant theory that worked well with the early synthesis experiments. Moreover, using a "locus theory" in the automatic generation of stop-vowel transitions in text-to-speech systems seems a plausible approach given the results obtained in intelligibility tests when the stimuli have been produced according to this model.

On the other hand, " locus equations " have been computed for various languages (Krull, 1987 for Swedish and Poch-Olivé, Fernández-Gutiérrez and Martínez-Daudén in this volume for Spanish and Catalan) showing the relationship between F2 at voicing onset following consonant release and F2 target vowel frequency; the plots obtained are assumed to indicate the degree of coarticulation. It may be pointed out that these equations have the same form and are based on the same parameters as those allowing the automatic derivation of CV transitions in text-to-speech systems.

Going back to the question of intelligibility, it is not difficult to predict that the manipulation of " locus equations " will lead to different results in the synthesis of stop-vowel combinations. The assumption underlying our approach is that varying degrees of coarticulation as expressed by different " locus equations " will produce synthesized syllables with different degrees of intelligibility and naturalness.

To test this hypothesis, we have carried out acoustic analysis of the following materials recorded by a male native speaker of Spanish:

(a) unvoiced stop-vowel syllables spoken in isolation.
(b) the same syllables in words embedded in carrier sentences.
(c) the same syllables taken from a corpus of spontaneous speech.

This allows us to gather data on different degrees of coarticulation - in terms of F2 onset *vs* F2 target plots for each place of articulation - that are related to three different speech styles, ranging from what has been called "laboratory speech" (condition a) to a good approximation to "spontaneous speech" (condition b).

The data obtained has been used as control parameters for an OVEIV digital formant synthesizer, and synthetic stimuli have been created reproducing the range of variation found in the three speaking styles.

On the other hand, we have concentrated on the importance of F2 starting point as an indicator of the degree of coarticulation. For that purpose we have created several series of stop-vowel combinations in which this is the only parameter to be modified, taking into account the values observed in the acoustic analysis of styles a, b and c. Two further conditions have been added in that case: (d) minimal coarticulation, with F2 starting at the upper limit of F1 as measured in a spectrogram, and (e): maximal coarticulation, with F2 starting at the same frequency as the vowel target.

Systematic perceptual tests are being made to assess both the intelligibility and naturalness of these stimuli, and it is hoped that their results will be presented at the conference. Some preliminary results indicate a slight preference for condition c, showing that the intelligibility of a stimulus may be favoured by using data from spontaneous speech as control parameters for the synthesizer.

References

ALLEN, J. (1985) " A Perspective on Man-Machine Communication by Speech", *Proceedings of the IEEE* 73,11: 1541-1550.

ALLEN, J.- HUNNICUTT, M.S.- KLATT, D.H. (1987) *From Text to Speech. The MITalk System* . Cambridge: Cambridge University Press.

DELATTRE, P.C. (1969) " Coarticulation and the Locus Theory ", *Studia Linguistica* 23: 1-26.

DELATTRE, P.C.- LIBERMAN, A.M.- COOPER, F.S. (1955) " Acoustic Loci and Transitional Cues for Consonants", *Journal of the Acoustical Society of America* 27,4: 769-773.

KRULL, D. (1987) "Second Formant Locus Patterns as a Measure of Consonant-Vowel Coarticulation ", *PERILUS* 5 (Fall 1986 - Spring 1987): 43-61.

PISONI, D.B.- NUSBAUM, H.C.- GREENE, B.G. (1985) " Perception of Synthetic Speech Generated by Rule ", *Proceedings of the IEEE* 73,11: 1165-1676.

COMPUTATIONAL MODELS OF SPEECH PERCEPTION

Dominic W. MASSARO
Program in Experimental Psychology
University of California, Santa Cruz
Santa Cruz, CA 95064 U. S. A.

Three classes of computational models of speech perception are described, evaluated, and tested within the paradigm of speech perception by ear and eye (1). The classes of models are interactive activation, connectionist models with hidden units, and process models. Experimental research has documented that both auditory and visual information influence speech perception in face-to-face communication. The perceiver evaluates and integrates continuous information from the audible and visible sources, and perceives the pattern that makes the best fit with this information. The three classes of models account for this result in different ways.

Interactive activation claims that the information from one source modifies the modality-specific processing of the information from the other source (4). That is, for example, the auditory information influences the quality of the visual information, and vice versa. Unconstrained models with hidden units assume that the inputs from both modalities activate and inhibit a layer of hidden units between input and output layers of units (6). All input units are connected to all hidden units and all hidden units are connected to all output units. The distinguishing feature of these two classes of models is that separate auditory and visual inputs interact immediately. The separate sources of information are not maintained independently of one another at some level of processing. Both interactive activation and unconstrained hidden unit models represent a more general class of nonindependence models.

The third model makes a distinction among three stages of processing: evaluation, integration, and decision (1). It is assumed that the auditory and visual sources of information are evaluated independently of one another. Thus, their representation at the evaluation stage maintains modality-specific independence. The two sources are integrated at the next stage, and the representation at this stage reflects the joint contribution of both sources. The final stage makes a decision depending on the task at hand. Perceptual recognition utilizes the outcome of integration, whereas discrimination can access the independent representations resulting from evaluation.

Interactive Activation Models

Interactive activation models are usually formulated within the context of a network of processing units. One of the most fundamental assumptions of interactivation models is the two-way activations among processing units. The TRACE model of speech perception (4), for example, is an interactive-activation model in which information processing occurs through excitatory and inhibitory interactions among a large number of simple processing units. These units are meant to represent the functional properties of neurons or neural networks. Three levels or sizes of units are used in TRACE: feature, phoneme, and word. Features activate phonemes which activate words, and activation of some units at a particular level inhibits other units at the same level. In addition, units at one level can activate and inhibit one another.

Interactive activation models correctly predicts that perceivers are influenced by both auditory and visual information in speech perception. The explanation of this phenomenon is that perceivers

have learned an association between the auditory and visual patterns of speech. This association is represented by interconnections among the auditory and visual units. In addition to predicting the joint influence of audible and visible information in speech perception, interactive-activation models predict a nonindependence between these sources of information.

The Roberts and Summerfield (5) study was formalized in terms of whether selective adaptation in auditory speech perception is purely auditory, but their results also speak to the issue of interactive activation of the auditory and visual information. In selective adaptation, listeners are exposed to a number of repetitions of an adapting syllable and then tested on a speech continuum between two speech categories. Relative to the baseline condition of no adaptation, the identification judgments of syllables along the speech continuum are pushed in the opposite direction of the adapting syllable. As an example, adaptation with the syllable /be/ (rhymes with *say*) decreases the number of /be/ judgments and increases the number of /de/ judgments along a /be/-/de/ synthetic speech continuum. Roberts and Summerfield (5) employed 7 different adaptors to evaluate the contribution of auditory and visual information to auditory adaptation along a /be/ to /de/ continuum. The adaptors were auditory /be/ and /de/, visual /be/ and /de/, audiovisual /be/ and /de/, and an auditory /be/ paired with a visual /ge/. (It should be noted that visual /ge/ and /de/ are essentially indistinguishable and similar results would be found with visual /de/ or visual /ge/.) After adaptation, subjects were tested on an auditory continuum between /be/ and /de/. If auditory and visual units interact, as assumed by interactive activation, then adaptation to a syllable in one modality should influence later processing of the syllable in the other modality. A lack of interactive activation would imply that auditory adaptation should be only a function of the auditory characteristics of the adaptor and independent of its visual characteristics.

Roberts and Summerfield found no evidence for interactive activation. Equivalent levels of adaptation were found for an auditory adaptor and a bimodal adaptor with the same phonetic information. The visual adaptors presented alone produced no adaptation along the auditory continuum. The most impressive result, however, was the adaptation obtained with the conflicting bimodal adaptor. The adaptor auditory /be/ paired with visual /ge/ produced adaptation equivalent to the auditory adaptor /be/. This result occurred even though the subjects usually experienced the bimodal adaptor as /de/. Thus, the adaptation followed the auditory information and was not influenced by the visual information or the phenomenal experience of the bimodal syllable. An interactive activation model would predict that the activation of the phoneme /d/ would provide top-down activation of the features representing that phoneme. Thus, subjects should not have adapted to the bimodal syllable experienced as /de/ in the same manner as their adaptation to an auditory syllable experienced as /be/. This falsification of interactive-activation models, such as the TRACE model, appears to be significant and not easily remedied because the interconnections among units appear to be central to the interactive activation models.

Network Models with Hidden Units

Models with hidden units are also formulated within the framework of network models. These models postulate a layer of hidden units between an input layer and an output layer of units. Unconstrained models with hidden units assume that all input units are connected to all hidden units and all hidden units are connected to all output units (6). In a theoretical and analytical analysis, I have shown that models with hidden units can be superpowerful—that is, they can predict many types of results and even results that do not occur (2). Not only are network models with hidden units too powerful, they appear to be wrong in the same way that interactive activation models are wrong. Hidden units that communicate with all input units introduce the same type of nonindependence effected by interactive activation. In both types of models, separate sources of information are not maintained independently of one another at some level of processing. Both interactive activation and unconstrained hidden unit models represent a more general class of nonindependence models. One

distinguishing prediction of these two classes of models is that separate auditory and visual inputs interact immediately in face-to-face speech perception. There is now a comprehensive body of results rejecting nonindependence in bimodal speech perception (1).

Process Model Assuming Independence Among Information Sources

A model assuming independence among the separate sources of information is the Fuzzy Logical Model of Perception (FLMP). According to the FLMP, speech patterns are recognized in accordance with a general algorithm (1, 3). The model assumes three operations in speech recognition: feature evaluation, feature integration, and decision. Continuously valued features are evaluated, integrated, and matched against prototype descriptions in memory, and an identification decision is made on the basis of the relative goodness of match of the stimulus information with the relevant prototype descriptions.

Testing between Independence and Nonindependence Models

The FLMP differs from interactive-activation models and unconstrained models with hidden units in terms of how multiple sources of information are brought together in speech perception. Although tests between the models might appear to be difficult, the two models make fundamentally different assumptions that can be tested by a fine-grained analysis of the predictions of the models against empirical results. The predictions of the two classes of models differ from one another when the models are conceptualized within the theory of signal detectability (TSD). A distinction is made between sensitivity and bias in TSD. Within this framework, the question is whether the top-down context modifies the sensitivity at the phoneme level, whether it modifies the bias, or both. That is, the question of interest is to what extent the top-down effects are reflected in sensitivity or bias (d' or β). The FLMP predicts no systematic effects of top-down context on sensitivity at the bottom-up level (3). In contrast to the predictions of the FLMP, both interactive activation models and models with unconstrained hidden units predict sensitivity differences. In interactive activation models, the top-down effects of phonological constraints can occur because of interactive activation between the word and phoneme levels. Bottom-up activation of the phonemes activates words, which in turn, activate the phonemes that make them up. Interactive activation appropriately describes this model because it is clearly an interaction between the two levels that is postulated. The amount of bottom-up activation influences the amount of top-down activation, which then modifies the bottom-up activation, and so on. A similar result occurs with unconstrained hidden-unit models because both bottom-up and top-down information influence the same hidden units.

Thus, an important empirical and theoretical question is whether top-down context produces sensitivity differences at a bottom-up level. This question was addressed by a study in which phonological context and segmental information were varied in a factorial design (3). A set of syllables along a /li/-/ri/ continuum was factorially combined with different initial consonant contexts /t/, /p/, or /s/. Subjects asked to recognize /l/ or /r/ were influenced not only by the information specifying /l/ or /r/, but also by the initial consonant context. For example, subjects reported /r/ more often in the context /t-i/ than in the context /s-i/. Without the appropriate experimental design and data analysis, we do not know if this result is due to sensitivity or bias. In this design, however, sensitivity effects would be reflected in changes in the discriminability of two adjacent levels along the liquid continuum as a function of context. Bias would be reflected in a change in overall response probability as a function of context. It should be noted that a bias effect is *not* artifactual or uninteresting; it can be as perceptual as a sensitivity effect. Sensitivity and bias are indexed by different dependent measures. Given that the subjects responded /l/ or /r/ on each trial, a measure of sensitivity is reflected in the differential responding to the two types of trials. To the extent the subject responded /l/ to one member of the adjacent pair and /r/ to the other member, sensitivity is high. Using the concept of information within information theory, the subjects transmits more information to the extent that there is differential

responding to the two stimulus alternatives. The overall probability of responding with a given alternative, independently of the stimulus that was presented, reflects bias within the framework of signal detection. To the extent the subject responds with only one response alternative, there is a bias towards that alternative. The results indicated a strong influence of phonological context on bias, but not on sensitivity. This result is strong evidence in favor of the FLMP over the class of nonindependence models, such as TRACE.

Discussion

From a decision-making perspective, interactive-activation models and unconstrained hidden-unit models are nonoptimal because they allow the processing system to distort the environmental input more than is reasonable. Given a movie review from two friends, discrepancies in the reviews lead to an eventual distortion of the original reviews within a model of interactive activation. The fact that one review differs from another, however, should not necessarily question the validity of either review. Given opposing reviews, the receiver of the reviews, however, might want to conclude very little about the value of the movie, but yet would be well-informed about each of the separate reviews. That is, the evaluation of each review is informative for the system but the integration leads to some ambiguity. In other cases, each of two sources will be somewhat ambiguous but consistent with one another. Integration in this situation provides more certainty than contained in either of the two sources. The integrity of the two reviews is preserved at the evaluation stage even though their integration provides more information than provided by each one separately. The stage representation of the FLMP allows for lower-level information to remain independent of higher-level information, although a *perceptual decision* about lower-level information will reflect the contribution of the higher-level information. In nonindependence models, however, the contribution of higher-level sources of information to lower-level decisions must come at the expense of modifying the representation of the lower-level information. The results seem to indicate that the representation of the lower-level information is *not* modified by higher levels, supporting the independence assumption of the FLMP.

References

1. MASSARO, D.W. *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale, NJ: Lawrence Erlbaum, 1987.

2. MASSARO, D.W. Some criticisms of connectionist models of human performance. *Journal of Memory and Language, 27,* 1988, 213-234.

3. MASSARO, D.W. Testing between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology,* in press.

4. McCLELLAND, J.L. -- ELMAN, J.L. The TRACE model of speech perception. *Cognitive Psychology, 18,* 1986, 1-86.

5. ROBERTS, M. -- SUMMERFIELD, Q.U Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *Perception & Psychophysics, 30,* 1981, 309-314.

6. RUMELHART, D.E. -- McCLELLAND, J.L. (Eds.) *Parallel distributed processing: Vol. 1. Foundations.* Cambridge: MIT Press., 1986.

ON THE ORIGIN OF THE SYMBOLIC VALUE OF SPEECH SOUNDS

Istvan T. MOLNAR
Institute of Slavic Philology
Kossuth University, Debrecen, Hungary

Introduction

Let's start with the following axiom: "Wherever we look, we find elemental phonetic symbolism" (8). There is no doubt about the validity of this statement. As has been proved by a great number of observations and experiments, speech sounds can evoke certain images and associations in the human mind. This problem has been the topic of linguistic investigation ever since Plato up to the present age. A new impetus to the problem was given in the modern age by Sapir's experiments (6) and Jespersen's observations (1, 2), who contributed substantially to the investigation of the semantic aspects of speech sounds. Of late there has been an abudance of experimental research of this problem (4, 9), but a number of unsolved questions still remain. One of these is the origin of sound symbolism.

Methods

Valuable contribution to the problem of origin has been provided by experiments aimed at the definition of the elemental symbolic value of Hungarian speech sounds. The experiment involving more than 900 persons as informants, included the whole spectrum of Hungarian speech sounds. Using Osgood's method for the measurement of meaning (5), we present the elemental values of Hungarian speech sounds with the help of the following oppositions: paeceful - fighting, pleasant - unpleasant, womenly - manly, merry - sad, slow - fast, soft - hard, strong - weak, light - heavy, hot - cold, small - large, beautiful - ugly, smooth - rough, dry - wet, rounded - angular, sweet - bitter. The data were processed and evaluated with the help of a computer.

Results

Our results are without any doubt indicative of Hungarian speech sounds having symbiloc values. Nearly half of the values of sounds symbolism measured in 16 sense-oppositions was found to be significant. We found vowels to be more "active" than consonants in producing effects of sound symbolism. Vowels were found to be more effective along the scales "bright - dark" and "merry - sad", i.e., they are connected primarily with visual stimuli

and the mood of the individual. Consonants, on the other hand, seem to be connected more with tactile sensations (soft - hard, hot - cold, smooth - rough), and with the notion of power and aggressiveness (peaceful- fighting, strong - weak).

We also investigeted the articulatory and acoustic features as well as the frequency of sounds, in addition to shape of the letters used to spell a particular sound.

Among the vowels the opposition betwen palatal/velar articulatory features was found to show a close correlation, closely followed by the opposotion betwen illabial/labial. These two oppositions are frequently in operation jointly, in a parallel fashion. So our "most merry", "brighest" vowel is palatal & illabial [i], whereas the "saddest", "darkest" one is velar & labial [u].

In the case of consonants a certain degree of correlation betwen the articulatory features and values of sound symbolism can be established. It is to be noted, however, that not only one, but several articulatory features in cooperation are operative in evoking a particular value of sound symbolism. Of particular interest in this respect are features connected with the place and the manner of articulation. We have found [r], [k], and [dz] to be our "hardest" consonants, and [h], [l], and [j] the "softest". The presence or absence of the "voice" element in articulation was also found to be important in this respect.

Regarding acoustic features, among the vowels the formant patterns, and in the domain of consonants the relation betwen the voice and noise elements as well as the sound being continuous or intermittent were found to be significant in producing different values in sound symbolism.

The correlation worth noting betwen the frequency of sounds and sound symbolism was only found in the correlation "beautiful - ugly". Here we can recognise the connection with our conventional notion of something being beautiful, i.e., pleasing our senses.

Our experiments have proved that the shape of letters has but little relevance to the symbolic value of sounds. The only exception, however, is provided by the sound [j], which, according to the rules of Hungarian spelling, is spelled either as "j", or "ly". The results is peculiar: our subjective appreciation of "ly" is much worse than that of "j". It seems to us that in this case we have a problem of transposition of our difficulties in spelling "j" or "ly" in our childhood to our personal feelings in adulthood concerning the sound symbolism of this consonant.

Discussion

We gained valuable results by the analysis of the results of our experiments as to the origin of sound symbolism.

There are two theoretical approaches to this problem. According to the first, sound symbolism is derived from the meaning of the word. E.g., in English, the sound [g] has the symbolic value "large", because it is often found in words like "great", "grand", "gargantuan", "gain", "gross", "grow" (7). Others, however, regard this theory of language habits as an incorrect interpretation of an otherwise correct observation (3). According to this view the cause and effect relation applies in an inverse manner, i.e., certain sounds, as a results of historical processes of language development, have been incorporated into words belonging to various semantic domains, but retaining their sound symbolism obtained earlier, from other sources. According to this view the source of sound symbolism is to be found in human voice itself, in the accompanying sensible and visible articulatory movements, as well as in the physical characteristics of the human sound, in the acoustically perceptible features.

The results of our investigations in the area of Hungarian speech sounds seem to support this latter view. On the basis of the correlations and various other relations established in the course of our investigations, we can summarise our results as the following: we think that the origin of the symbolic value of human speech sounds is to be found in the process whereby the various articulatory, perceptual and motor elements in the production of speech sounds are, in some peculiar way, transferred into the appropriate areas of the human mind. In view of this we advance the following hypotheis: the human voice is perceived by our mind as something of natural origin, as a natural attribute of the world that we live in. In this way speech sounds can acquire semantic content. Hungarian [i], for example, is closely associated with the following semantic features: "little", "bright", "fast", "light", "merry", "womanly", "beautiful", "pleasant", "sweet", whereas [k] is regarded as being close to "hard", "angular", and "strong".

Alongside the articulatory and acoustic features as primary sources of sound symbolism, other sources should also be considered. Our investigation show that such are the frequency of sounds and the shape of letters. The effect of these however is of much less inportance.

Only after a thorough investigation of the sources can we answer the question concerning sound symbolism: is it language-specific or a language universal? As sound symbolism has been found characteristic of every language so far investigated, it should be regarded a language universal. As regards the sound symbolism connected with individual sounds, we can expect parallelism among languages depending on the similarity or identity of the articulatory or acoustic features of the sounds concerned. As is known, betwen sounds of different languages that are known to be similar, various difference in articulation and acoustics can be demonstrated, therefore there seems little doubt that

this fact is concomitant with diverse values in the sound symbolism of the sounds in question.

References

1. JESPERSEN, O.: Sound symbolism. In: Language: its Nature, Development, and Origin. London, Allen and Unwin, 1922, 396--411.
2. JESPERSEN, O.: Symbolic value of the vowel [i]. In: Linguistica. Copenhagen, Levin and Munksgaard, 1933, 238--303.
3. LEVICKIJ, V.V.: Semantika i fonetika (Semantics and phonetics). Chernovtsi, Chernovtsi University, 1973.
4. MIRON, M.S.: A cross-linguistic investigation of phonetic symbolism. Journal of Abnormal and Social Psychology 62. 1961, 623--30.
5. OSGOOD, C. E.-SUCI, G. J.-TANNENBAUM, P. H.: The measurement of meaning. Urbana, University of Illions Press, 1957.
6. SAPIR, E.: A study in phonetic symbolism. Journal of Experimental Psychology 12. 1929, 225--39.
7. TAYLOR, I. K.: Phonetic symbolism re-examined. Psychological Bulletin 60. 1963, 200--9.
8. TAYLOR, I. K.-TAYLOR, M. M.: Another look at phonetic symbolism. Psychological Bulletin 64. 1965, 413--27.
9. ZHURAVLEV, A. P.: Foneticheskoe znachenie (Phonetical meaning). Leningrad, Leningrad University, 1974.

PHONE RESTORATION

Bruno H. REPP
Haskins Laboratories
270 Crown Street, New Haven, CT 06511-6695, USA

Introduction

In the original demonstration of the "phonemic restoration illusion" (4), subjects listened to a sentence in which the acoustic signal pertaining to one phoneme, the fricative /s/, had been excised and replaced with an extraneous sound (a cough) or with silence. When the cough was present, the subjects claimed to hear the speech as intact and could not localize the cough accurately within the sentence. The silence, however, was usually heard as replacing the /s/. Later research (e.g., 2,3,5) extended the finding to other phonetic segments and other extraneous sounds, and established that the relationship between their acoustic properties affects the strength of phonemic restoration.

In theory, there are two ways in which such restoration could take place. One possibility is that it occurs entirely at the phonological level. In that case, the extraneous sound replacing the missing phoneme provides merely "bottom-up confirmation" (3) of a restoration that has already taken place on the basis of contextual and lexical information. The other possibility is that the missing speech is fashioned from the acoustic material provided by the extraneous sound. If so, the speech is not really missing but hidden, and the listener's task is to separate it from that portion of the extraneous sound which cannot be incorporated into the speech stream.

These two possibilities have implications for the perception of the extraneous sound, which has not been examined in previous studies. According to the first hypothesis (the phonological restoration hypothesis), the sound replacing part of the speech signal in a sentence should be perceived more or less as it would be perceived in isolation. According to the second hypothesis (the auditory restoration hypothesis), the sound should be perceptually impoverished relative to its occurrence out of context; that is, it should be perceived as missing that part of its energy and spectral structure which characterizes the restored speech signal.

The purpose of the present research was to test these hypotheses. The task required listeners to pay attention to a target noise which replaced just the fricative noise of /s/ in a sentence (hence, "phone restoration") and to compare its auditory quality to a probe noise occurring either before or after the sentence. According to the auditory restoration hypothesis, the target noise should sound lower in pitch or brightness than an identical probe noise; according to the phonological restoration hypothesis, they should sound the same. Since embedding a noise in a sentence may exert an independent effect on the perceived timbre of the noise, a control condition was included in which the extraneous noise was inserted into the silent closure interval of a /t/. Any change in the subjective pitch of the embedded noise in that condition may be attributed to the sentence context per se; any additional change in the /s/ condition must then be due to auditory "subtraction" of the restored phone from the target noise.

Experiment 1

The stimuli were created by editing digitized waveforms of speech produced by a female speaker. The target noise occurred either at the beginning or at the end of a trisyllabic word embedded in the constant carrier phrase "Say ... again". The probe noise occurred either before or after that phrase. The critical words in which the target noise replaced the fricative noise of /s/ were "seminar" and "happiness". The control words containing /t/ were "telephone" and "cabinet". These words were repeated many times in the course of the experiment, but it was hoped that this would not eliminate the phone restoration effect.

Two noise targets (referred to as NS and NSS) were paired with three noise probes (N, NS, and NSS). N was created by low-pass filtering a burst of white noise having the same amplitude envelope as the original /s/ noise (S); it had the lowest pitch. NS was produced by waveform addition of N and S; it had a nearly flat spectrum and an intermediate pitch. NSS resulted from waveform addition of NS and S; it had the highest pitch. According to the auditory restoration hypothesis, NS (NSS) targets should sound like N (NS) probes after an /s/ noise has been restored.

Pairings of N with NSS were omitted, as they were too easy to discriminate. Thus there were five target-probe combinations. Each of these occurred in four different speech contexts, with the probe either preceding or following the speech, for a total of 40 trials that were presented to 10 subjects 8 times in random order. A pretest with isolated noise pairs preceded the main test. The subjects indicated whether the second noise sounded higher, equal, or lower in pitch than the first noise.

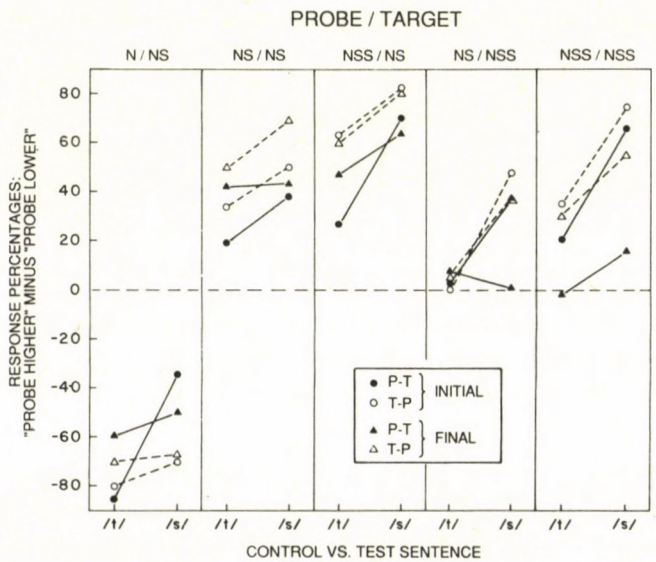


Figure 1.

The results are shown in Figure 1. Each of the five panels compares the results for control (/t/) and test (/s/) trials for one of the five target-probe combinations. The four functions in each panel represent the two possible temporal orders of target and probe (P-T or T-P) for

each of the two possible target locations in the test word (initial or final). The quantity plotted on the ordinate is the difference between the percentage of judgments that the probe had a higher pitch than the target, and the percentage of opposite judgments. These scores thus range from -100 to 100, with 0 representing perceived identity. "Equal" judgments were not considered.

The most striking feature of Figure 1, and also the most important result, is that in nearly all conditions the /s/ trial probes received higher scores than the /t/ trial probes, which means that target noises sounded lower in pitch when they replaced the frication of /s/ than when they replaced the silence of /t/. This result supports the auditory restoration hypothesis. The corresponding main effect in an ANOVA was highly significant, $F(1,9) = 23.83$, $p = 0.0009$. Yet, NS (NSS) targets were not perceived as identical to N (NS) probes, which suggests that phone restoration was incomplete. There was also a general bias to judge probes as higher in pitch than targets.

Experiment 2

Experiment 2 partially replicated Experiment 1 with probes of a finer grain. A continuum of five probe noises (N1-N5) was constructed by mixing the waveforms of N (=N1) and NS (=N5) in various proportions. Two of them (N3 and N5) were used as targets. Because of the increase in probe-target combinations, only the initial-location stimuli ("seminar" and "telephone") were used. Ten subjects listened to 10 randomized blocks of 40 trials each, following practice with pairs of isolated noises.

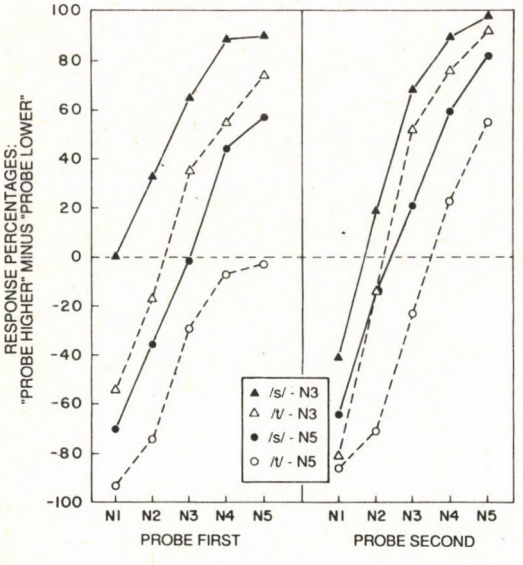


Figure 2.

The results are shown in Figure 2. On the abscissa we have now a continuum of probe noises, with separate panels for preceding and following probes. It is evident that subjects' responses increased as an orderly function of probe pitch. Probe discrimination was better when the probe followed the speech, as indicated by the steeper functions.

Clearly, the two targets were also discriminated, with higher ratings for probes paired with N3 than for probes paired with N5. Most importantly, there was a consistent difference between the /s/ and /t/ conditions: For all probe-target combinations, the probe was judged higher (hence, the target was judged lower) when /s/ was to be restored, in agreement with the auditory restoration hypothesis. The effect was highly significant, $F(1,9) = 83.27$, $p < .0001$.

As in Experiment 1, however, /s/ restoration was far from complete. If it had been, the N5 target should have been judged as matching the N1 probe; in fact, N5 was judged closer to N3 or perhaps N2 (filled circles in Figure 2). A large part of that shift was in fact due to the general bias to perceive probes as higher than the target, which was again present: In three out of four pairings of probes with identical targets replacing /t/, the probe was judged as higher in pitch than the target. (Only when the N5 probe preceded the N5 target were they judged to be the same in pitch, but judging from the general shapes of the functions, this result seems anomalous.) Thus the /s/ restoration effect, though consistently present, is relatively small.

Conclusions

These two experiments demonstrate that perceptual restoration of a fricative noise involves the subtraction of spectral energy from the noise replacing it. Part of that noise is perceived as a fricative (though perhaps not a perfect one), and the remainder is perceived as an extraneous sound. Since the restoration was apparently not complete even when the /s/ noise was physically present in the target noise, these data also show that listeners have difficulty separating an extraneous noise from fricative noise if the two coincide exactly (cf. 1).

Several questions are left unanswered by these preliminary experiments: (1) Can the same effects be obtained with more varied stimulus materials and/or with different target phones? (2) Is the phone restoration effect induced by residual phonetic cues in the signal (e.g., formant transitions) or does it have a true top-down (i.e., lexical) component, like the classical phoneme restoration effect? (3) What is the origin of the bias to hear probes as higher in pitch than targets? (Did listeners perhaps "restore" /s/ on control trials as well?) Further experiments are in progress which address these questions.

References

1. DARWIN, C.J.: Auditory processing and speech perception. In H. BOUMA and D.G. BOUWHUIS (Eds.), Attention and Performance X. Hillsdale, NJ: Erlbaum, 1984, 197--209.
2. SAMUEL, A.G.: Phonemic restoration: insights from a new methodology. Journal of Experimental Psychology: General 110. 1981, 474--94.
3. SAMUEL, A.G.: The role of bottom-up confirmation in the phonemic restoration illusion. Journal of Experimental Psychology: Human Perception and Performance 7. 1981, 1124--31.
4. WARREN, R.M.: Perceptual restoration of missing speech sounds. Science 167. 1970, 392--3.
5. WARREN, R.M.--OBUSEK, C.: Speech perception and phonemic restorations. Perception & Psychophysics 9. 1971, 358--63.

(This research was supported by NICHD Grant HD01994.)

SPEECH PERCEPTION IN CHILDHOOD AND ADULTHOOD: CONTINUITY OR DISCONTINUITY?

Natalie WATERSON
Department of Phonetics and Linguistics, University of London, London, England.

It is now widely acknowledged that there is continuity from babbling to speech as against the earlier view that there was a clear break between the two. Now there is controversy as to whether there is continuity between the acquisition of the first 50 words or so, which are acquired slowly, and the speech development that follows, which is characterized by a more ordered phonology and a rapid expansion in vocabulary.

When child phonology began to be the object of serious study, the emphasis was on production and children were assumed to acquire their mother tongue by the learning of phonemes rather than whole words. In the 1970s, with more attention being focused on perception, the view was put forward that children learn whole words from the start. This led to a divergence of opinion: 1) that children start and continue to learn whole words rather than segments - the continuity view, and 2) that children start the learning process with words, for the first 50 words, and then change to phonemes - the discontinuity view.

The view that children change to phonemic processing seems to have arisen from the need to describe their speech processing in line with what was the current understanding of adult speech processing rather than from any findings of child language research. Now, after decades of acoustic research, it has still not proved possible to find invariant acoustic cues for phonemes which would justify their use as units of speech perception, invariance being necessary to explain normalization across variables of different speakers, age, sex, and accent. Larger units are therefore being considered as more suitable candidates (2). There is also a wider choice of phonological theories available now and phonemes are not the only option. Thus the main justification for considering that children change from word to phoneme processing as language development progresses no longer exists.

We may ask ourselves why the phoneme was accepted uncritically as an appropriate unit for speech processing in adult speech research in the first place. It seems it was due to a historical accident: it was not arrived at as a result of research in speech processing as would be the normal expectation but was due to the fact that phoneme theory was the only available theory that acoustics phoneticians could turn to in the early days. Phoneme theory developed under the influence of the alphabetic system of writing; the I.P.A. (International Phonetic Alphabet) is based on the alphabet, and transcriptions and orthographies show words as visually presented sequences of segments (letters), so it was logical to assume that words are heard as auditorily presented sequences of segments (phonemes). The use of phonemes as descriptive units in child language research followed naturally as phoneme theory was still the only theory available at the time and the same symbols were used to transcribe child and adult speech. Phoneme theory inevitably led to the assumption that children learn phonemic contrasts in the acquisition of the phonology of their mother tongue, and this became the major focus of research. The change to the distinctive feature approach made little difference as the learning of distinctive features still involved the acquisition of phonemic contrasts. However it is a long time before a child meets minimal pairs which provide him with a direct means of learning phonemic contrasts and it may be questioned whether there is indeed any incontrovertible evidence that the acquisition of phonology proceeds by way of learning phonemic contrasts. What evidence is there that is not open to other interpretations? Child forms like [dɔ] or [gɔg] for 'dog' [dɔg], [du:] for 'Juice' [dʒu:s], [bʌp] for 'bump' [bʌmp], [ɪʃ] for 'fish' when represented in transcription may suggest phonemic learning - they are represented as sequences of segments - but it is equally possible to interpret them as examples of whole word learning.

into account in the child's construction of his forms and structures. This was a regular phenomenon of his early phonological development in relation to other consonants and also to vowels, viz., responding first mostly to features that are acoustically and auditorily salient and then paying attention to those that are less salient gradually over time.

In the adult forms considered above, the vowel of the first syllable differed but in each case it was more open than the vowel of the second syllable:

		More open	More close
Adult	'Bobby'	[bɔ]	[bɪ]
	'birdie'	[bɜ:]	[dɪ]
	'bucket'	[bʌ]	[kɪt]
	'button'	[bʌ]	[tn̩]
	'Patrick'	[pæ]	[tr:k]

and the child's [bæbu:] had the same relationship of vowels, viz., more open: [bæ] more close: [bu:] . Although the degree of openness of vowel is responded to accurately (Acoustic cue: F1 (First formant) which has relatively high intensity), the features frontness, backness and centrality of the first syllable are not so accurately produced, nor are the features of frontness and spreading of the second syllable (Acoustic cues: F2 and F3, which have relatively less intensity and are hence less salient than F1). Instead there is backness and rounding in the child's second syllable which seems to require less precision in articulation after a bilabial consonant than frontness and spreading.

In cases where the adult forms had the same degree of vowel openness in both syllables, e.g., close-mid in 'biscuit' [bɪskɪt] and 'pudding' [pʊdɪŋ], the child also had the same degree of vowel openness in both syllables : [be:be:] 'biscuit', and [pʊpʊ] 'pudding'. Here he had PVPV where the adult had PVSPVP for 'biscuit' and PVPVN for 'pudding'. This sort of relationship of plosive to plosive and degree of openness of vowel was a regular feature of the child's forms at this age, 1;6 ((6) p.87) and suggests that the child was recognizing the same pattern in words of the type described above.

The data of the child P were typical of English children in many ways; the analysis showed that he responded in a regular patterned way to sets of adult words which share certain features. From this it may be deduced that children recognize patterns in adult words on the basis of features that are salient for them, i.e., the most acoustically and auditorily salient ones, and abstract such features from the acoustic signal as cues for the construction of their own structures and patterns but within the constraints of their production and organizational limitations. Similar patterned relationships to those shown in the Waterson studies can be found in the data of other children (1) (3) (4). Thus children may be said to attend selectively to certain features of a word, to the auditory pattern of a whole word, a skeleton spectrum rather than to a full spectrum which they then have to segment. Such auditory patterns, based on a limited number of acoustic cues, can be shown to be invariant, as described below, and are proposed as possible candidates for units of speech perception for both children and adults. Detailed discussion of this is given in Waterson (6).

The cues of a pattern are in a particular relationship and in a particular temporal sequence - cues such as, for instance, peaks of intensity for number of syllables; duration: some syllables longer, some shorter, and intensity: F1 for the relationship of degree of openness of vowel, e.g., the vowel of one syllable is more open than that of the next, or more close, or the same; Fo (Fundamental frequency) for intonation; and the different cues marking the major classes of consonant, showing, for instance, syntagmatic contrasts such as nasal stop at the onset of one syllable, oral stop in the next, and so on. The cues are relative to each other so such relationships are invariant and hence the pattern is invariant; the actual frequencies, actual degrees of intensity, actual durations, actual consonant and vowel qualities are variable from speaker to speaker and vary with rate of speaking but do not affect the pattern as the same relationships hold. Such patterns are invariant across variables of age, sex, different speakers, accents, and rates of speaking, and can thus offer an explanation of how child and adult normalize: how they see samenesses where there are acoustic differences. The way the invariant auditory pattern is used in the perception and interpretation of continuous speech is described in Waterson (5) and (6).

The fact that children may respond to some but not all features of a phoneme needs explaining: if some features from a phoneme bundle are missing, is it still the same phoneme? It is known that in children's perception and production, classes of sounds (e.g., plosive, nasal, fricative, etc.) are recognized and produced correctly before contrasts of place (e.g., bilabial, alveolar, velar, etc.) within the classes. This can mean that there is recognition but non-production of place because of production problems but it can equally well indicate that speech sounds are not perceived and responded to as bundles of features, cf. the findings in adult speech that adults perceive features independently (8) (9).

Children often have a number of 'words' which are based on what seem to be unanalysed chunks of adult speech (formulas), like [issæ] for 'What's that?' [wɔts ðæt]. If these are to be explained in phonemic terms, there are several complex operations implicit which could well be beyond the capacity of a young child. Even such cases as quoted by Weir ((7) pp.108-9) as an example of a 'conscious exercise of learning a new phonemic contrast', can be given another interpretation. Weir's son was contrasting two variations on the first vowel of 'berries': [bɛriz] not [bɜrɪz] - [bɜrrz, bɜrɪz] - not [bɜrɪz] - [bɛri:z]. It is possible that rather than learning a new phonemic contrast, he was working on the whole word, aware that he was not saying it correctly and was trying to get it right. He did not have 'Barry' [bɜrɪ] or 'Barry's' [bɜrɪz] in his repertoire so the two forms were not minimal pairs for him. In fact he had three variants for adult's [æ]: [æ], [ɛ], and [e] and was possibly impressing on his mind which was the correct one for the particular word.

A telling argument against the phoneme segment as a unit in perception and inner representation in children is their inability to segment speech until they learn the alphabet and to read and write. Experiments have shown that children who can talk quite well and who are able to process adult speech at the normal adult speaking rate, are unable to segment speech into its component sounds, and this has been found to apply to children up to the age of seven. Some interesting experiments with Russian children were conducted by Zhurova (10). She tried to get children of 3-4 years, 5-6 years and 6-7 years to separate out the first sound of words and found that they had considerable difficulty. The only way she could get them to do so was by prolonging the initial sound of the word first herself, e.g., [m-m-midɥed] 'ɯɐ̯ɔ̯ɬɐɔ̯ɓ' = [b-b-bɔːs] 'bear'. Even the 6-7 age group was found to be incapable of identifying the individual sounds in words. Zhurova's experiments showed that children have to be trained to isolate sounds in a word. It seems highly unlikely that there could be any processing or internal representation of words in terms of segments if children are not able to segment words into separate sounds.

It is when research turned to the study of earlier vocalizations, and when hardly recognizable attempts were included in the studies, and more detailed phonetic transcriptions began to be made, that it became evident that a phonemic analysis was not able to handle many of the aspects of child phonology. Analysis in terms of whole words made it possible to show that early, hardly recognizable words have a holistic, not a segment to segment relation to adult words (6). This opened up the possibility that easily recognizable words too may have such a relation to the adult model and that learning involves pattern recognition (6). This, combined with an appreciation of the young child's physiological and cognitive immaturity and the fact that perception develops gradually, made it possible to give simple explanations for differences between child and adult forms, and between child earlier and later forms. For instance, the child P (6), at age 1;6, had homonyms for 'Bobby' [bɔbɪ], 'birdie' [bɜːdɪ], 'bucket' [bʌkɪt], 'button' [bʌtn], and 'Patrick' [pætrɪk]: all these were [bɜbu] for him. The adult forms had CVCV or CVCVC structures, C_1 and C_2 were plosives (P) and C_3 where present, was a plosive (P)or a nasal (N): PVPVP and PVPVN. The child responded with CVCV and PVPV to all five adult forms. He did not have a final C. Why did he not respond to the final consonants of the adult forms? It seems this was because he was responding to what was auditorily salient for him and was not giving much attention to what was less salient. Plosives at the beginning of a syllable are acoustically and auditorily more salient than those at the end of a word, which are weakly articulated; the same goes for nasals, and these weak non-salient final consonants were not taken

An invariant auditory pattern of the type described above would make for economy in speech processing time as the number of cues involved is very small, so could account for the rapidity of speech processing within the appropriate contexts; variability in the acoustic shape of a word, unless extreme, would not cause problems of interpretation.

Child language research has shown that children rely heavily on non-linguistic sources of information for their understanding of speech: their limited capabilities and simple level of language mean that they will at first make less use of the acoustic signal of speech than adults and thus use a minimum of acoustic cues. As shown above, their auditory patterns will be much simpler than those of adults and will be based on a more limited number of salient cues - on a skeleton spectrum of the adult form. The skeleton is fleshed out within the constraints of what the child is capable of producing and his phonological organization. Similar patterns have been proposed as units in adult speech perception (6), but adults use more cues and thus have more complex patterns than children. As pointed out above, in the course of language acquisition, children gradually make use of more and more of the less salient cues so their patterns eventually match those of adults.

The hypothesis presented here claims that children begin to process speech in the way they will continue to do when adults, that there is continuity in the processes of speech perception and recognition throughout speech development and that this can be described as speech processing by pattern recognition. The adoption of the concept of an invariant auditory word pattern as a unit of speech perception may well provide new insights for the study of phonological development as well as for speech processing by adults.

References.

1. LEOPOLD, W.F. : Speech Development of a Bilingual Child 2. Evanston, North Western University Press. 1947.
2. PERKELL, J.S. & KLATT, D.H. (eds.) Invariance and Variability in Speech Processes. Hillsdale, Laurence Erlbaum. 1986.
3. SMITH, N.V. : The Acquisition of Phonology. Cambridge, Cambridge University Press. 1973.
4. VELTEN, H.V. : The growth of phonemic and lexical patterns in infant language. Language 19, 281-92. 1943.
5. WATERSON, N. : The role of patterns in language acquisition. In E. Oksaar (ed.) Sociocultural Perspectives of Multilingualism and Language Acquisition. Tubingen, Gunter Narr. 1987.
6. WATERSON, N. : Prosodic Phonology: the Theory and its Application to Language Acquisition and Speech Processing Newcastle, Grevatt & Grevatt. 1987.
7. WEIR, R.H. : Language in the Crib. The Hague, Mouton.
8. WICKELGREN, W. : Distinctive features and errors in short-term memory for English vowels. Journal of the Acoustical Society of America 38. 1965.
9. WICKELGREN, W. Distinctive features and errors in short-term memory for English consonants. Journal of the Acoustical Society of America 39. 1966.
10. ZHUROVA, L.Ye. : The Development of analysis of words into their sounds by preschool children. In C.A. Ferguson & D.I. Slobin (eds.) Studies of Child Language Development. 141-54. New York, Holt, Rinehart & Winston. 1973.

AUTOMATIC DETERMINATION OF ALLOPHONES

Katarina BARTKOVA & Denis JOUVET
Centre National d'Etudes des Télécommunications
Route de Trégastel
22300 LANNION - FRANCE

Introduction

Allophones, unlike phonemes, provide a more detailed description of the possible acoustical realizations of the sounds. The introduction of allophones in a speech recognition system based on a Markov modelling approach [2], allows to take into account the contextual influence, and improves the recognition performances. But the determination of the allophones, which is generally done using a spectrographic representation of the sounds [1], can be very long and requires expert knowledge. On the other hand, expert-determined phonetic cues cannot always be detected automatically.

Thus, in developing an automatic procedure for the determination of the allophones that uses the same acoustical parameters as the speech recognition system we are provided with allophones that match the distinction capabilities of the speech recognition system. This paper presents such a procedure which is based on an automatic classification of the acoustical realizations of the sounds. As the procedure relies on an acoustical representation of the right and left contexts of the sounds, we first tested the allophones resulting from a manual segmentation and labelling of the data base, then those resulting from an automatic segmentation. In both cases, on a data base of French numbers, the introduction of these allophones improves the speech recognition performances. Finally, the classification procedure was applied to study the influence of the sound duration on its acoustical realization.

Description of the data base

The data base used in this study contains about 5300 french numbers, between 0 and 999, recorded from 70 speakers (45 men and 25 women). This data base was divided into two almost equal parts : a training set and a testing set. The training set, containing the data recorded from 34 speakers, was used to determine the automatic allophones and the "optimal" parameters of the Markov model. The testing set, containing the data recorded from the 36 other speakers (about 2700 numbers), was used to obtain speaker independent speech recognition performances.

The whole data base was recorded in a quiet room and digitalized at a sampling frequency of 12.8 kHz. The data of the training set have been hand-segmented and labelled. After the acoustical analysis, each word (or sentence) of the data base is represented as a sequence of vectors of 6 Mel frequency cepstral coefficients. These vectors are computed every 20 ms (frame rate) using the energy in 24 Mel filters.

Automatic determination of the allophones

The procedure for the automatic determination of the allophones uses a representation of the sounds as a sequence of three prototypes (vector quantization) and classifies the acoustical realizations of the sounds corresponding to the different contexts.

Representation of the sounds In order to reduce the amount of data to manipulate, each sound occurring in the data base is coded as a sequence of 3 spectral frames representing the left context, the target, and the right context. These frames usually correspond to the first, the middle and the last frame of the speech segment. However, when the duration of the sound is greater than 60 ms (more than 3 frames), the target spectra is obtained by averaging the 2 or 3 middle frames. A vector quantization based on a Lloyd and splitting algorithm [3] is applied to the spectral frames (vectors of 6 cepstral coefficients). Each sound is then represented by its duration and a sequence of 3 prototypes (quantization of the 3 spectral frames).

Method of classification Using all the sounds corresponding to a same phoneme in a given context, we computed the normalized histograms of occurrences of the quantization prototypes for the target and the left and right contexts. In the study of the sound duration influence, the histograms are computed for different segment durations (for example 50 to 70 ms, 70 to 90 ms, and so on). The distance between two acoustical realizations of a given sound has been defined as the average distance between the histograms corresponding to the left contexts, the targets and the right contexts. Using a distance threshold we grouped the acoustical realizations into classes. The lower this threshold will be, the more classes we will obtain, and a large threshold will yield only one class per sound.

Discussion of the results In a previous study [1], using a spectrographic representation of the sounds, we proposed, for the same data base, an average of 2.1 allophones per vowel and 2.7 allophones per consonant. The automatic procedure seems to be more sensitive to the vocalic variations : for classification thresholds ranging from 30 to 60, the number of allophones varies from 3.5 to 1.6 per vowel and from 3.3 to 1.3 per consonant. The differences between the "manual" allophones and the "automatic" allophones are not uniform.

For example, with the "manual" allophones, we considered only one acoustical realization for the schwa-like vowel regardless of the surrounding consonantal contexts. However, as the main suprasegmental feature of this vowel is a short segment duration, its spectral variation could be great : for 9 different contexts, our algorithm gave us 8 allophones for the schwa with a classification threshold of 30, and 5 allophones with a threshold of 60. On the contrary, for the closure of plosive /k/ as well as for the closure of /t/, the differentiation between a nasal and an oral vowel left context was found with a classification threshold of 30. But when this threshold increased, the allophone groups became less homogeneous and this differentiation disappeared leading to only one allophone, whereas it was one of our main allophone features in the previous study. Also, concerning the /t/ and /k/ burst, the automatic method gave less importance to the distinction between the bursts followed by a pause and those not followed by a pause, as well as to the burst palatalization (when followed by the labio-palatal /Y/).

Recognition performances using the allophones

To validate this automatic determination of the allophones we introduced them into a speaker independent speech recognition system and tested the recognition performances on the French numbers between 0 and 999.

Description of the speech recognition system The speech recognition system is based on a Markov modelling approach. The system uses gaussian probability density functions that are associated to an acoustical network. The "optimal" value of the model parameters (the probability of the transitions and the gaussian parameters) is automatically determined by maximizing the emission probability of the training data. The single integrated network corresponds to all the possible sentences of the application and is obtained by compiling all the a priori knowledge of the application : syntax, lexical description of the words, phonological rules, and specification of the acoustical models for the basic units (phonemes or allophones). The use of a compiled network and phonetic basic units enables the handling of coarticulations either by allowing alternate pronunciations or by differentiating the basic units depending on the context, thus introducing allophones.

We used this last feature of the network compiler to introduce the phonological rules that reflect the results of the automatic classification of the acoustical realizations. The rules were introduced in such a way that they defined, for each phoneme, as many allophones (basic units of the model) as the number of classes of the automatic classification procedure. Each allophone reflects the acoustical realizations of the phoneme in the contexts occurring in the corresponding class. However, some theoretical contexts might not be specified in the rules if they did not occur in the labelled data used for the automatic classification. Thus in order to avoid these problems

some a priori classes of contexts were defined by grouping the phonemes having a similar influence : for example, the fricatives /s/ and /z/ were grouped together as they have the same acoustical influence on the adjacent vowel.

Allophones resulting from a manual segmentation Several classifications of the acoustical realizations were made and the recognition tests were performed for thresholds ranging from 30 to 60. The following table shows the recognition error rate on the entire numbers, and the last column corresponds to a simple model using a phonetic description without any allophone.

Table 1 - Recognition error rate on the french numbers using phonological rules resulting from different classification thresholds and ***a manual segmentation***.

Classification threshold	30	40	50	60	---
Number of gaussian pdf	425	321	253	227	149
Error rate on training set	5.6 %	5.6 %	6.8 %	6.9 %	6.9 %
Error rate on testing set	6.5 %	6.1 %	7.3 %	7.5 %	8.1 %

Its appears from the table above, that a small classification threshold gives a large model (i.e. many gaussian pdf) and that the performances increase with the size of the model. The introduction of phonological rules improves the acoustical description of the application, even if the results are not as good as with the "expert-defined" allophones. Some differences between the models were due to the necessity of introducing phonological rules that exactly reflect the results of the automatic classification procedure (for example, clusters not present in the labelling were not used).

Allophones resulting from an automatic segmentation In the previous tests, the allophones were obtained using a manual segmentation and labelling of the training data. Such a task being long and tedious we would like to avoid it. So, using a simple Markov model, without any allophone, we carried out an automatic segmentation and labelling of the training data, by finding the most likely path in the network that corresponds to any given sentence. Then applying the same procedure we tested the recognition performances for different classification thresholds ranging from 20 to 50.

Table 2 - Recognition error rate on the french numbers using phonological rules resulting from different classification thresholds and ***an automatic segmentation***.

Classification threshold	20	30	40	50	---
Number of gaussian pdf	540	431	284	238	149
Error rate on training set	4.7 %	5.4 %	6.0 %	6.2 %	6.9 %
Error rate on testing set	6.3 %	6.6 %	6.9 %	6.8 %	8.1 %

We note a similar behaviour of the models : as the threshold decreases, the size of the model increases and so does the performances. The results obtained from this automatic segmentation, which are comparable to those obtained from a manual segmentation, are quite encouraging for an extension of the procedure to a larger data base.

Duration influence of the sound on its acoustical realization

By computing the normalized histograms using only the segments having a specified duration (for example 30 to 50 ms, 50 to 70 ms, etc) it became possible to apply the same classification procedure to study the influence of the sound duration on its acoustical realization. We first took into account only the histograms resulting from the middle (target) frames, thus ignoring the contexts.

French, unlike English or Swedish, does not have a heavy accent, and non-stressed or weekly stressed vowels will not approach a schwa in their quality. Nevertheless, in french, the accent is strongly correlated with the sound duration. The sounds occurring in a stressed position (the last syllable of a rhythm group) are always longer than those occurring in a non-stressed syllable. Thus, the probability of reaching the articulatory target position is greater for the sounds occurring in a stressed syllable than elsewhere. This tendency appeared obviously in our data. We were able, not only to notice the correlation of the sound duration and its acoustical realization, but also to detect the duration threshold needed to reach the target position for vowels as well as for consonants.

We observed that beyond a certain segmental duration, the central sound spectral frames were grouped into the most homogeneous classes by the vector quantization algorithm. For our data, this crucial duration (threshold) was around 90 ms. Out of a total of 5276 vowels segments, with a duration greater than 90 ms, 97.8% of the middle (target) vectors were coded by the most homogeneous (89%) and the second most homogeneous classes (8.8%). For the consonants, among the 5685 occurrences whose duration reached the crucial 90 ms, 97.9% of the middle (target) vectors were coded by the most homogeneous (94%) and the second most homogeneous classes (3.9%). This proves that the possible interspeaker or contextual variations in the steady-state sound position, beyond a given duration value, are irrelevant for an automatic coding system. These results were independent of the number of coding classes as well as of the mean distortion value.

When we tried to apply the same procedure to the left and right context histograms, we did not notice a similar behaviour. The coding fluctuations were always great. Also two facts are worth mentioning : first, the accurate detection of the sound boundaries is rather difficult, and second, as the spectral analysis was carried out at a 20 ms frame rate, the first and last frames of the segments were not always centered on the boundaries.

According to our observations, it seems obvious that below a sound duration threshold, using a phoneme model in an automatic speech recognition system is not reasonable. The differences between the acoustical realizations of a sound in a given context due to the sound duration could be as important as the spectral variations due to different contexts. Thus, using the acoustical information of a sound without using its duration might not yield reliable results.

Conclusion

Though the results obtained in this study are hopeful, it appears evident, that an automatic method of research of allophones requires a very large data base, containing enough occurrences of every context to provide reliable normalized histograms. If a context has only a few occurrences, its corresponding histogram will not be reliable, and included in a class, it can corrupt the whole grouping. That is the reason why we should implement a mixed method : on the one hand we need to introduce phonetical rules in the speech recognition system (or in the classification procedure) in order to deal with the unusual or missing contexts (interpolation or averaging with other data), and on the other hand we can automatically determine allophones for the well represented contexts.

References

[1] K. Bartkova, D. Jouvet : "Speaker-independent speech-recognition using allophones"; Proc. ICPhS 1987, Tallin, USSR, pp. 244-247, August 1987.

[2] D. Jouvet, J. Monné, D. Dubois : "A new network-based speaker-independent connected-word recognition system"; IEEE Proc. ICASSP 1986, Tokyo, pp. 1109-1112, April 1986.

[3] H-Y. Su : "Reconnaissance Acoustico-Phonétique en Parole Continue par Quantification Vectorielle"; Thèse de doctorat, Rennes, 1987.

USE OF PHONETIC FEATURES IN A SPEECH RECOGNITION SYSTEM BASED ON HIDDEN MARKOV MODELLING

Dominique DUBOIS, Guy MERCIER
Centre National D'Etudes des Télécommunications
Route de Trégastel, 22301 Lannion, FRANCE

INTRODUCTION

In our laboratory, two different speech recognition algorithms have been developed. The first one, based on statistical methods, is used in a speaker independent connected word recognition system named Phil 86. It uses a cepstral analysis carried out every 20 ms. The other one, based on an analytical method, is used in a speaker dependent continuous speech recognition system named Keal. It uses an acoustic-phonetic analysis based on spectral analysis, segmentation and phonetic labelling.

The goal of this work was to use the Keal acoustic-phonetic module as a preprocessor for the statistical recognition system Phil 86. The phonetic segmentation decreases the number of parameters and therefore reduces the computation. We have tested this approach for two different outputs of the phonetic analysis of Keal, the acoustical coefficients and the phonetic features, and compared the performances with cepstral coefficients computed on each phonetic segment.

After a brief description of the Phil 86 system, we shall detail the acoustic-phonetic decoder of Keal and we shall present these experiments.

DESCRIPTION OF THE STATISTICAL SYSTEM, PHIL 86

The Phil 86 software [1, 2] enables speaker-independent speech recognition. It is based on a hidden Markov modelling of the words to be recognized. The acoustical analysis is a Mel Frequency Cepstral Analysis (MFCC). The result of this analysis is an 8-dimensional vector combining total energy and cepstral coefficients, computed every 20 ms.

The Markov model is defined by a set of states, a set of transitions and a set of acoustical distributions. The states and the transitions define the structure of the "acoustical" network, and the distributions, associated to the transitions, define the probabilities of emitting the acoustical spectrum. Each distribution, represented by a gaussian probability density function (pdf) with a diagonal covariance matrix, is entirely defined by the mean and standard deviation of the acoustical coefficients.

The integrated acoustical network, representing all the possible sentences of the application, is compiled from all the a priori knowledge of the application : syntax, lexical description of the words, phonological rules and specification of the acoustical basic models (phonemes). The "optimal" set of parameters (probability of the transitions and gaussian parameters) is computed in order to maximize the probability of emitting all the observations of the training set. The entire application being modelled by a single acoustical network, the recognition algorithm, using Viterbi decoding, is limited to look for the path providing the best match with the sequence of frames computed from the unknown sentence.

PRESENTATION OF THE ANALYTICAL SYSTEM, KEAL.

Keal is a hierarchical bottom-up speech recognition system [3], which combines statistical, structural and knowledge-based pattern recognition techniques. An unknown utterance is recognized by means of the following procedures: acoustical analysis, phonetical segmentation and identification, word and sentence analysis. The task to be performed, described by its vocabulary and its context-free grammar, is given as a parameter of the system.

The acoustic-phonetic decoder is based on a set of speaker-independent deductive rules, able to segment speech signals into phones and to recognize the main coarse phonetic features characterizing these segments. In parallel, a statistical decision based on a set of linear discriminant functions allows to refine these first phonetic hypotheses. A speaker adaptation module computes the parameters of these linear discriminant functions by matching known utterances with their acoustical representation.

The lexical module is mainly based on a statistical dynamic programming technique which performs a matching between a phonemic lexical entry containing various phonological forms and a phonetic lattice. The sentence recognition algorithm is derived from Earley's parser. The functionality of this module is to reconstruct the uttered sentence as a complete string of words, starting from the lexical lattice obtained at the previous step.

Acoustic-phonetic decoding in Keal

This module starts by a sentence onset detection based on several criteria (averaged energy, localisation of vowels ...) that progressively refines the precise detection of the speech limits. Each speech frame is then labelled as "consonant", "vowel" or "silence" using a set of rules and acoustical cues derived from the basic parameters. The sentence is segmented into pseudo-syllables using mainly the search for the syllable vowel nucleus. A segmentation into phones is then realized. From this segmentation, the extraction of phonetic features described below and the phone segments identification are performed to obtain a phonetic lattice which is the main output of the acoustic-phonetic module. The list of the possible phonemes is ordered with decreasing likelihood. Let us briefly describe this segmentation which we will use in our experiment.

Segmentation into phones and evaluation

First of all, vowels are located around the maximum of energy inside each syllable. Sequences of stationary and transient events are then located between each vocalic nucleus. The main phonetic features are identified in these segments. The rate of compression brought by the segmentation is about 10.

An evaluation of this segmentation [4] was carried out for 7 speakers on 39 sentences of the "pseudo-logo" language in an application of vocal programming. The data base contained 5601 phonemes, each sentence having an average length of 20 phonemes.

We call *omission* the fact that a segment has been eliminated in the segmentation process, and *insertion* the introduction of a superfluous one which should be connected to one of its neighbours. On the above data, the percentage of omission was of 3.5% and the percentage of insertion was of 10.5%. One has to consider that it is in general admitted that omissions are worse errors than oversegmentation ; a closer analysis shows that an important proportion of insertions and omissions come from the beginning and end of sentences ; this is a problem concerning the silence/speech separation. Such insertions are often recognized as fricative consonants or voiceless plosives.

Description of the acoustical coefficients

Acoustical analysis is carried out by an 14-channel vocoder and the acoustical spectrum is computed every 13.3 ms. The frequency bandwidth of each channel is logarithmically distributed from 250 Hz to 4300 Hz and the frequency bandwidth covered by the first 12 channels corresponds to the telephone bandwidth. Additional parameters like voiced-unvoiced decision, pitch, signal amplitude, spectral centers of gravity are also measured and used by the phonetic recognizer. After the "vocalic-non vocalic" decision, a vector of 29 coefficients is defined for each frame. These coefficients are then averaged over every phone segment. The acoustical vector contains 13 coefficients representing the differences of energies between consecutive vocoder-channels and 16 other coefficients: the center of gravity of the total energy distribution, the energies defined inside four frequency sub-bands, the center of

gravity of the energy distribution in these sub-bands, the local temporal variations of the total energy and of the center of gravity between the current frame and the preceding one, the maximum jump of energy on all the frequency bandwidth or after preaccentuation on the vocoder-channels, some differences between some maxima of vocoder-channels.
This vector of 29 acoustical coefficients (for each segment) will be used as input to the Phil 86 system in the experiment EXP2.

Definition and extraction of phonetics features

A degree of plausibility of each phonetic feature, varying from 0 to 100 percent, is computed on each segment. The basic parameters for the computation are the energy in several frequency bandwidths, the flatness of the spectrum, F0, etc...

The consonantic segments are characterised by seven features: plosive, fricative, voiced/unvoiced, nasal, liquid, velar and dental. The voiced/unvoiced feature is detected by the F0 measurement, the energies in the 250-650 Hz bandwidth and the ratio between low and high frequency energies. The fricative feature is detected by the same energy ratio, the spectral center of gravity and the degree of flatness of a spectrum.

The vocalic segments are characterised by three features: open/closed, front/back and oral/nasal. The two first features are computed at three places in the vowel segments (beginning, middle, end). The open/closed feature, for instance, is detected by one cue based on comparison of energies computed in the low frequencies, less than 1050 Hz. These frequency ranges are tuned according to the mean pitch value. This cue is computed on three selected frames after the vocalic nuclei have been located and a degree of openness is assigned to these frames.

To the fourteen coefficients representing these features we added the energy of the segment and a weighting factor ; therefore we obtain a vector of sixteen coefficients representing the phonetic features which constitutes another input set to the Phil 86 system in the experiment EXP1.

EXPERIMENTS

In order to test the influence of the segmentation of Keal on the Phil 86 system, we used, on one hand, the phonetic features and the acoustical coefficients of Keal and, on the other hand, cepstral coefficients as they are computed in Phil 86, on each segment determined by Keal. The advantage of the Keal segmentation is that it reduces the volume of data to be treated by the markovian system. The goal of the following experiments is to determine whether such a reduction of the computational load does not deteriorate the recognition rates.

In our experiments we use a set of parameters computed on each phonetic segment of Keal on which we apply a Viterbi decoding in a markovian network (Phil 86). Three experiments have been pursued:

EXP1: associating the Keal Segmentation with a vector of 16 **phonetic features**
EXP2: associating the Keal Segmentation with a vector of 29 **acoustical coefficients**
EXP3: associating the Keal Segmentation with a vector of 14 **MFCC based coefficients.**

This last experiment (EXP3) uses the same coefficients as provided by the acoustical analysis of Phil 86 system, i.e. the energy and 6 cepstral coefficients for the middle frame of the segment, and 7 other coefficients representing their variation between the beginning and the end of the segment.

The data base used for the recognition test contains digits pronounced by 64 speakers and numbers between 0 and 999 pronounced by 70 speakers. The training uses the data from half of the speakers (32 for isolated digit, 34 for the numbers) and the recognition test are realized on the other half (32 other speakers for the digits and 36 other speakers for the numbers). Therefore the results that we give are speaker independent recognition performances. The size of the data base is about 500 digits for the training set and 500 for the

test set and more than 2500 numbers for the learning set and 2500 for the test set. The digits and the numbers have been used separately and represent the two data bases treated. We give only the **recognition error rate** on the complete numbers, first on the training set then on the test set. As a comparison, we recall the results of the Keal and the Phil 86 systems on these same data bases.

Coefficients		Digits Data base		Numbers Data base	
Type	Number	Training	Test	Training	Test
EXP1 (phonetic features)	16	5.8 %	11.0 %	39.0 %	40.0 %
EXP2 (acoustical coefficients)	29	0.8 %	4.5 %	24.0 %	26.0 %
EXP3 (MFCC)	14	1.6 %	3.0 %	12.7 %	13.0 %
KEAL		5.4 %	5.8 %	17.9 %	18 %
PHIL 86		1 %	1 %	4.9 %	4.3 %

These experiments show very clearly that some sets of coefficients are not suitable to characterize the speech signal, in particular, the set of phonetic features which gives twice more errors than the Keal system and ... ten times more than the Phil 86 system. The gaussian modelling with covariance diagonal matrix may be ill-suited to the manipulation of the phonetic features or the acoustical coefficients of Keal. The association of MFCC and of the segmentation is better than the Keal system alone but always less good than the Phil 86 system. Thus, it is obvious in this table that the a priori segmentation degrades the Phil 86 performances (three times the error rate). Indeed, the segmentation reduces the computation but it introduces errors which can't be corrected by the Markov system. The degradation of the recognition performances is too important to justify its use.

CONCLUSION

As the phonetic segmentation introduces errors not corrected by the markovian system, one should tries some other kind of a priori segmentation. For example, a compression of the acoustical data based on the signal stability might reduces the computation load without introducing the same problems.

An other point worth mentioning is the coherence between the gaussian pdf and the MFCC coefficients. However, this kind of pdf might not be well adapted to model the phonetic features. Thus, it would be interesting to study the behaviour of such a system (involving phonetic features combined with a statistical approach) using discrete probability density functions.

REFERENCES

[1] JOUVET D., MONNE J., DUBOIS D. : *A new network-based speaker-independent connected-word recognition system.* ICASSP Tokyo, p.1109, 1986.

[2] JOUVET D. : *Reconnaissance de mots connectés indépendamment du locuteur par des méthodes statistiques.* Thèse ENST Juin 1988.

[3] MERCIER G., BIGORGNE D., MICLET L., LEGUENNEC L., QUERRE L. : *Recognition of speaker-dependent continuous speech with KEAL.* To appear in I.E.E. Proceedings, 1989.

[4] MICLET L., MERCIER G. : *Evaluation of the acoustic decoder of the KEAL Speech Recognition system.* Proceedings of 9th I.C.P.R., Rome, Italy, 1988.

METHODS FOR DECREASING THE RESPONSE TIME IN ISOLATED WORD SPEECH RECOGNITION

András FARAGÓ, Géza GORDOS, Gábor LUGOSI
Speech Research Laboratory,
Institute of Communication Electronics,
Technical University of Budapest, Hungary

1. Introduction

This paper places the focus on methods that have been proved to be efficient in speeding up the response time of the speaker dependent isolated word recognition system VERBIDENT, developed in 1987 [1].

As a measure of performance of such a system we suggest the following factor:

$$\frac{(\text{response time}) \times (\text{error probability})}{\text{size of vocabulary}}$$

the reason to define such a compound measure of performance is that it is desirable to exclude pseudo-solutions, which decrease the response time simply by giving up a large part of recognition accuracy or by decreasing the vocabulary size, without really improving the overall performance.

As our main method for speeding up the recognizer we present a sophisticated Nearest Neighbor algorithm, which makes possible to find the vocabulary element that matches optimally the input word, by evaluating only a small fraction of all possible DTW (or other) distances. This algorithm is called Geometric Search, as it constitutes, in some sense, a geometric analogue of the well-known binary search procedure.

The algorithm, as implemented in the system VERBIDENT, gave an improvement in performance of about *one order of magnitude*, measured in the above mentioned factor of performance.

Now we describe the algorithm in a general form, since it is applicable not only in speech recognition but in other fields as well. (For a more detailed description and analysis see [2].)

2. The algorithm

Denote the training points (prototype words in the vocabulary) by $t_1, t_2, \ldots, t_k$ and the point to be classified (the word to be recognized) by x. They are all elements of a metric space M.

The algorithm is based on two rules:

a.) Exclusion rule

Let t be a training point for which the distance $d(x,t)$ is still unknown and t^* be another training point for which the distance $d(x,t^*)$ has been already evaluated during the decision process. Furthermore, denote by r_{min} the minimum of all those distances, which have been already evaluated. Then, by the triangle inequality, t can not be the nearest neighbor of x if one of the following inequalities holds:

$$d(x,t)+r_{min} < d(t,t^*)$$

$$d(x,t)-r_{min} > d(t,t^*)$$

In this case the point t can be excluded from further considerations, without evaluating $d(x,t)$.

Remark: Experimental results show that a considerable proportion of training points can be excluded in this way from distance evaluation, which is profitable with respect to CPU-time, because computing the DTW metric is complicated, hence excluding takes far less time then distance evaluating. The price for this is that we have to calculate and store all the distances between training points. But this can be done in a preprocessing stage, not affecting the real-time capabilities.

b.) Selection rule

Suppose the the distances $d(x,t_{i_1}), d(x,t_{i_2}), \ldots, d(x,t_{i_k})$ have been already evaluated. Let T be the set of those training points for which the distance $d(x,t)$ has not yet been calculated and which have not been excluded by a previous application of *Exclusion Rule*. Then, for the next distance calculation select the prototype $t \in T$ for which the following function is minimum:

$$f_k(t) = \sum_{j=1}^{k} |d(t_{i_j},t) - d(x,t_{i_j})|$$

(If the minimizing point is not unique, choose the one with smallest index.)

Remark: The idea behind the *Selection Rule* is the following: if the training points are distributed in the spase densely enough, then for any training point t we have:

$$d(x,t) \approx d(t^{NN},t),$$

where t^{NN} stands for the nearest neighbor of x. Therefore, the training point which minimizes $f_k(t)$ is a prospective candidate for being a nearest neighbor of x.

Now we describe our proposed algorithm:

Algorithm "Geometric Search":

Step 0. (Initialization) Set $t := t_1$, $t^{NN} := t_1$, $r_{min} := \infty$, $T = \{t_2, \ldots t_k\}$

Step 1. Calculate $d(x,t)$

Step 2. If $d(x,t) < r_{min}$ then $r_{min} := d(x,t)$ and $t^{NN} := t$

Step 3. Apply the *Exclusion Rule* to update T (delete t and all the exluded points from T).

Step 4. If T is empty then STOP, the last value of t^{NN} is the nearest neighbor and its distance from x is r_{min}. If T is not empty then go to Step 5.

Step 5. Apply the *Selection Rule* to select a new $t \in T$.
Return to Step 1.

Informally, the idea of Geometric Search can be explained as follows. Suppose we have evaluated the distance $d(x,t_1)$. Then, if the sample points are distributed densely enough, $d(t^{NN},t_1) \approx d(x,t_1)$ holds, where t^{NN} is the nearest neighbor of x. That is, ideally the further search could be restricted to the surface of a ball of radius $d(x,t_1)$, centered at t_1. The Selection Rule reflects the intention to choose the next point as close to this ball surface as possible. Call the next point t_2. Then, by the same reasoning, we can restrict the search to the intersection of two ball surfaces of radius $d(x,t_1)$ and $d(x,t_2)$, centered at t_1 and t_2, respectively. Proceeding further, the search is restricted to the intersection of more and more spheres. Thus, one can expect that the number of training points, which are still to be examined, decreases exponentially, showing an analoguous behavior to the binary search procedure.

Of course, the above reasoning works only in the asymptotic sense. With a finite sample set we cannot expect the nearest neighbor to be positioned exactly on some spheres. Therefore, in the algorithm we exclude only the points, which are surely not nearest neighbors (by the Exclusion Rule) and choose the next point on the principle that it is "collectively" close to all already known spheres (by minimizing the selector function) .

References

[1] Faragó A., Gordos G., Koutny I., Magyar G., Osváth L., Takács Gy.: "A VERBIDENT-SD-2 izolált szavas gépi beszédfelismerő", Hiradastechnika, 1988/3, pp 111-116.

[2] A. Faragó, T. Linder, G. Lugosi, T.Pikler,: "Geometric Search: A fast exact nearest neighbor algorithm", submitted to IEEE Trans. on Pattern Analysis and Machine Intelligence.

REVIEW OF SOME OF THE ACTIVITIES AT THE SPEECH RESEARCH LABORATORY OF THE TECHNICAL UNIVERSITY OF BUDAPEST

Géza GORDOS
Speech Research Laboratory
Technical University of Budapest, Hungary

Introduction

The paper reviews some of the activities of the Speech Research Laboratory of the Technical University of Budapest in an approximately chronological order.

Modeling pitch perception

In the model developed in 1974 it was assumed that pitch percepction by human could be modeled by spectral analysis [4]. In case of a voiced (quasi-periodic) signal the spectrum depends very much on the length of the segment of signal selected for analysis. Namely, the decay of the spectrum with increasing frequency is the quickest if the length of the segment equals the pitch period or its multiple. Consequently pitch determination can be formulated as finding that length of segment for which the decay of the spectrum is the quickest.

Classification of Identical and Fraternal Twins by Speech Processing

Distinguishing between identical and fraternal twins, i.e. between monozygots and dizygots, is indispensable in genetics. Blood tests are cumbersome and require much blood. New method [3] uses appr. 1 minute speech recordings by each person. Fourteen feature parameters are extracted from each recording. Two of the fourteen parameters are new findings. A vector is formed from the two sets of fourteen parameters belonging to a particular pair of twins. Using a multi-variable discriminance analysis the vector is classified as either monozygotic or dizygotic. No decision error was found on a 17-member test population. (Research still in progress.) The method is very useful in speaker verification, too.

Speech Detection in Severe Noise

A method aimed at the automatic detection of human speech in very noisy environment have been worked out [5]. The speech--to-background noise ratio can be as low as -6dB. Detection is based on number of zero crossings, spectral nonstationarity (i.e. the fluctuation of short time energy between various frequency bands) and a newly found property of the average magnitude difference function.

Speech Synthesis

A very high quality text-to-speech (TTS) synthesis system for Hungarian and Experanto language have been developed in cooperation with the Phonetics Laboratory, Hungarian Academy of Sciences [8,9]. The system is based on a novel method of defining speech reproduction units. Attempts are being made to apply the method for other languages.

Limited vocabulary speech synthesis (LVSS) systems have also been developed on both LPC and formant coding principles. The word editing system for the LPC-PARCOR type synthesizer is 99% automatic and provides a 2 to 3 kbit/sec encoded rate [10]. A screen interactive word editing system for the formant type synthesizer provides less than 1 kbit/sec encoded speech [7]. This word editor is implemented on a board that plugs into an IBM-PC. Tailoring the system for different languages is the target for further development.

LVSS and TTS systems are implemented in various forms (e.g. attechment for Commodore'64; IBM-PC-board, etc.)

Speech Recognition

A speaker dependent, isolated word recognition system has been developed [1]. Recognition time is appr. 60 msec per item of vocabulary on a TMS 32010 processor. Speed is further increased by an original technique colled geometric search [2]. Error is less than appr. 2% with only one training pass required. Accuracy improves with increased number of training passes. The hardware is a printed board that plugs into an IBM-PC or compatible. Software for both vocabulary editing and recognition is user friendly. The recognition procedure is based on Dynamic Time Warping using data provided by PARCOR and zero crossing analysis. Replacing Dynamic Time Warping by an improved version of the Hidden Markov Model approach is in progress.

Speaker independent recognition is also in progress.

Recognition and text-to-speech systems have been integrated to form two-way-speech man-machine interface (1988).

Speech Enhancement

A speech enhancement system is based on adaptiv Wiener filtering. Signal and noise parameters are extracted from the composite signal and separated from each other by using an algorithm identifying pauses of speech.

Assessement of the Quality of Digitized Speech

An improvement on the MNRU (Modulated Noise Reference Unit) measuring method, recommended by CCITT to assess the quality of speech after decoding, has been implemented and tested. The basic idea is the use of additive noise, too, with power carefully related to that of the multiplicative noise [11].

Basic Research

The notion of complex formants, interpreted in the complex frequency domain, leads to improvements at the acoustic level of speech recognition as well as in speech synthesis. Complex formants are defined as the poles of the function

$$H(s)=\left\{1-\sum_{i=1}^{p} a_i e^{-s_i}\right\}^{-1},$$

where a_i's are the LPC coefficients of signal segments formed by periodic extention of one quasi period. VC and CV transitions can be analysed in this way in great detail.

Suggesting adaptive AMDF and the alternate use of forward and backward AMDF [6] are further examples of current basic research and find applications, among others, in improved voiced/unvoiced decision as well as pitch determination.

References

[1] Farago,A.,Gordos,G.,Koutny,I.,Magyar,G.,Osváth,L.: VERBIDENT: An Isolated Word Recognizer, 9th Colloquium on Acoustics, Budapest, 1988.

[2] Faragó,A.,Gordos,G.,Lugosi G.: Methods for Decreasing the Response Time in Isolated Word Speech Recognition, Proc. of Speech Research '89, Budapest.

[3] Forrai,G.,Gordos,G.: A new acoustic method for the discrimination of monozygotic and dizygotic twins, Acta Paediatrica Hungarica, Vol. 24, No.4., 1983. pp. 315-321.

[4] Földváry,R.,Gordos,G.: A Hypothetical Model for Pitch Perception by Human (in Hungarian) Hiradástechnika, Vol. XXV. (1974), No. 11., pp. 344-348.

[5] G.Gordos: "Speech detection in severe noise", Proc. of the 11th International Congress on Acoustics, Paris, 19-27. July, 1983. Proc. of... Vol.4. pp. 91-94.

[6] Gordos,G.: Prospects and Limitations in Speech Processing - Overview and Some Novel Methods, Proceedings of the 9th conference on Acoustics, Budapest, May 4-7, 1988.

[7] Gordos,G.,Németh,G.,Olaszy,G.,Tihanyi,A.: Embedding Speech Synthesis into Applications, Proc. of Speech Research '89, Budapest.

[8] Olaszy,G.,Gordos,G.: On the Speaking Module of an Automatic Reading Machine, First European Conference on Speech Technology, Edinburgh, 1988, Proc. pp. 25-28.

[9] Olaszy,G.: Speech Synthesis in Hungary from the Beginnings up to 1989, Proc. of Speech Research '89, Budapest.

[10] Podoletz,Gy.,Békési,S.,Gordos,G.,Takács,Gy.: LIAVOX, the First Hungarian Language Independent Speech Synthesis System, XVIII. Conference on Acoustics, Ceske Budajovice, 1985. (Proceedings)

[11] Tatai,P.: Comments on Objective Quality Measures in Speech Encoding, Royal Institute of Technology, Stockholm, Internal Rep.: IR-TTT-8802, 1988.

SOME RESEARCH ON PHONETICALLY BASED ISOLATED WORD RECOGNITION

GORDOS Géza, KOUTNY Ilona, OSVÁTH László
Institute of Communication Electronics
Technical University of Budapest

1. Introduction

In the recognition of isolated words, pattern recognition is widespread. For the recognition of a large vocabulary and especially of continuous speech this method is not convenient. Units smaller than words are needed: phones (phonemes), diphones or elements even smaller than one phone (transemes). The mapping of acoustic events into phonetic events is the bottleneck of this approach. Because of the relatively low accuracy of phonetic recognition, it must be combined with higher level means such as syntactic analysis.

Our research is directed to phonetic segmentation and identifying of Hungarian vowels in order to cope with large vocabulary speaker dependent isolated word recognition. The system works in the same framework as the small vocabulary isolated word recognizer Verbident (see [2]); i.e. the programs written in Pascal run on an IBM PC provided with a digital signal processing board based on TMS 320. The research is subsidised from the National Scientific Research Fund.

2. Feature extraction

The speech signal is sampled with 10 Khz every 15 ms. Feature extraction both by filtering and linear predictive coefficients has already been investigated in Verbident. Segmentation based on filtering is carried out at the Acoustic Laboratory of HAS [5] collaborating with us. Our research relies on LPC based formant tracking.

Starting from the LPC-s, partial correlation (PARCOR) coefficients of 15 ms length intervals of the speech signal are calculated. The PARCOR data may give a kind of short time spectrum estimation. Besides the first 10 coefficients the total energy (on a logarithmic scale) and two further pieces of data characterise the frames.

As is known, the i-th PARCOR coefficient reflects the correlation between samples separated by i steps in such a way that their virtual correlation, carried by the samples lying between them, is removed. All the PARCOR coefficients are normalised (a kind of correlation factor), so after the feature extraction there is no need to normalise them.

Several different methods are used to determine the PARCOR coefficients. The procedure chosen by us determines the samples of the correlation function from frame to frame for the whole word in the first step, from them it calculates the PARCOR coefficients afterwards [3]. This special kind of spectral

analysis is used to determine the formants.

The change of all these parameters is very quick for phone transitions; this is expressed by the parameter PD:

$$PD = \prod_{i=1}^{10} (1-k_i)^2$$ where k_i is the i-th PARCOR coefficient

This allows us to distinguish the pure and transitional phase of vowels.

3. Formant tracking

The first three formants are calculated for every frame, that is the three most intensive peaks in the frequency domain are determined. For this purpose the poles of an all-pole filter built from the PARCOR coefficients are used which is considered as a model for speech generation. The roots of the polinomial of 10-th degree are calculated by the method of Graeffe. Afterwards the roots which could be formants are selected from the maximum 10 roots.

The spectrogram of the word *beszédkutatás* [speech research] illustrates the complexity of the task. The program tries to find the same characteristics.

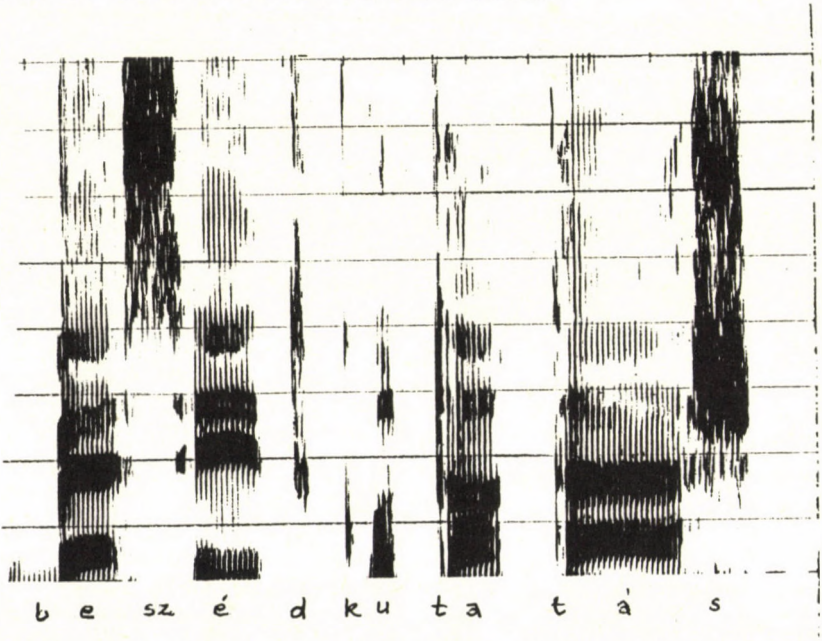


Fig. 1: Spectrogram **beszédkutatás** [speech research]

4. Looking for phonetic cues

Segmenting and labelling of speech is difficult because of the big interspeaker variability and the large contextual

variability in the acoustic structure of phonetic segments, just as it is also for human spectrogram readers [4]. So our first aim was to find the vowels, as they will serve as phonetic cues in the recognition. For this the local peaks in the time-energy function and the PD parameter are used. At the proposed places the 3 formants are compared to the reference patterns.

Several formant patterns for 9 Hungarian vowels (a: ɒ o u ø y ɛ e: i:) are stored in a speaker dependent reference dictionary. The long vowels are not considered separately, though they have a slightly different structure. The duration will be taken into consideration afterwards.

At present, the parameters of vowels are determined manually from a short recording of the speaker uttering a phonetically balanced word set. There are vowel prototypes from a starting position, after fricatives, plosives, nasals, etc., because of the distortion due to coarticulation. Stressed and unstressed variants are also included.

Furthermore some conjectures are made concerning the number and the class of consonants between the vowels. Therefore this rough phonetic labelling concentrates on the vowels which constitute the nuclei of the speech.

5. Word recognition

CV structure is fundamental in the natural languages. Dogil [1] calls them pivots and proposes a phonetic parser on this basis. So we have the skeleton of the word: e.g. CVCV, where the vowels are already known and there are some conjectures for the consonants. We are investigating which sounds can fill these slots also on a statistical basis.

A special algorithm tries to match this sequence to the words of the dictionary which is ordered according to the vowel structure. The nearest word will be the recognized word. This dictionary also contains the phonetic transcription of the words.

References

[1] Dogil, G. (1987): The Pivot Parser. In: Speech Technology. Ed. J. Laver & M.A. Jack. Edinburgh. p.112-115

[2] Faragó A., Gordos G., Koutny I., Magyar G., Osváth László (1988): VERBIDENT: An Isolated Word Recognizer. In: 9th Acoustic Conf. Ed. T. Pritz. Budapest. p.115-119.

[3] LeRoux J., Gueguen (1977): A fixed point computation of partial correlation coefficients. In ASSP-25 1977 N3

[4] Lonchamp, F. (1987): Reading spectrograms: The view from the expert. In: Fundamentals in Computer Understanding: Speech and Vision. Ed. J.-P. Haton. Cambridge: Cambridge University Press

[5] Vicsi K., M. Mattila, Berényi P.(1988): Continuous speech recognition using different methods. Internal report

AUDITIVE UNTERSUCHUNGEN ZUM SPRACHSYNTHESESYSTEM FLEX-DEUTSCH

Hans GRASSEGGER
Abteilung für Phonetik des Instituts für Sprachwissenschaft
Karl-Franzens-Universität, Graz, Austria

Einleitung

Jede Entwicklung eines qualitativ hochwertigen text-to-speech (im Folgenden: TTS) Synthesesystems bedarf einer systematischen Evaluation der mit ihm synthetisierten Sprache. Ein komplettes TTS-System besteht freilich aus einer Vielzahl von Komponenten aus Text-, Sprach- und Signalverarbeitung. Jede Verbesserung einer der Komponenten (z.B. Graphem-Phonem-Konversion, Intonationsalgorithmus) kann die Leistung des gesamten Systems erhöhen. Die Evaluation der Syntheseergebnisse darf sich nicht darin erschöpfen, lediglich in einem abschließenden Test den erreichten Standard zu demonstrieren. Vielmehr muß sie bereits während der Entwicklung eines TTS-Systems regelmäßig durchgeführt werden, damit die diagnostische Analyse der Resultate eines Evaluationstest zu entsprechenden Modifikationen des Systems führen kann, wie das etwa in einem niederländischen Forschungsprogramm vorgesehen ist (POLS, 1988a).

Die meisten Untersuchungen zur Evaluation von synthetisierter Sprache wurden bislang auf der segmentalen Ebene durchgeführt, obwohl die Notwendigkeit einer befriedigenden Synthese suprasegmentaler Charakteristika unbestritten ist (POLS, 1988b:24; TERKEN/LEMEER, 1988). Wenn unsere auditiven Untersuchungen zum TTS-System FLEX-DEUTSCH dennoch ebenfalls die segmentale Verständlichkeit zum Gegenstand haben, dann vor allem aus jenem Grund, den POLS (1988b:23) so treffend formuliert:

"None of the presently available rule synthesizers (...) have such a good segmental quality that one could further neglect this level and concentrate completely on higher level processing. All present systems will gain speech quality by improving segmental intelligibility."

Das TTS-System FLEX-DEUTSCH

Das deutsche TTS-System, welches durch unsere auditiven Untersuchungen evaluiert werden sollte, heißt FLEX-DEUTSCH und wurde am Linguistischen Institut (Phonetik-Labor) der Ungarischen Akademie der Wissenschaften in Zusammenarbeit mit der Technischen Universität Budapest entwickelt. Das System besteht aus dem frei programmierbaren Sprachsynthetisator MEA 8000 (Philips, 1983) mit drei Formanten (F1-F3), vier Bandweiten (B1-B4), Amplitude (AM), Zeitintervall (FD), Grundtonerhöhung (PI) und Grundton beim Start (Fo) als Steuerparameter. Auf nähere Details zum Synthetisator und zum Graphem-Phonem-Konversionsprogramm muß hier aus Platzmangel verzichtet werden, doch sei auf die ausführlichen Darstellungen bei OLASZY (1988) und bei OLASZY/GORDOS (1987; hier wird das mit gleicher Grundkonzeption entwickelte ungarische TTS-System SCRIPTOVOX vorgestellt) verwiesen.

Testmaterial und Testdurchführung

In einem ersten Schritt sollte die Verständlichkeit der mit dem TTS-

System FLEX-DEUTSCH generierten Konsonanten in initialer und intervokalischer Umgebung getestet werden. Dabei wurde von folgendem deutschen Konsonanteninventar ausgegangen: p,t,k,b,d,g,f,s,sch,ch,v,j,z,tsch,m,n,ng,r,l. Die Notation der Konsonanten in dieser Aufzählung ist bewußt eine graphemische (d. h.'z' ist nicht etwa [z], sondern [ts]), da ja die entsprechenden Laute bei orthographischer Eingabe über die Computer-Tastatur durch das Graphem-Phonem-Konversionsprogramm generiert werden sollen. Alle oben angeführten 19 Konsonanten wurden in intervokalischer Position (VCV) getestet, für die initiale Position (CV) war aus phonotaktischen Gründen 'ng' auszuklammern und auf anlautendes 'ch' wurde wegen seiner Beschränkung auf Fremdwörter und des gelegentlichen Schwankens seiner Realisierung als [ç] oder [k] verzichtet, so daß im Anlaut nur 17 Konsonanten getestet wurden. Der aufmerksame Leser wird im obigen Konsonanteninventar zwei Elemente vermissen, nämlich 'pf' und 'h'. Die Entscheidung, diese beiden Elemente (vorläufig) noch nicht zu testen, sondern erst in einem späteren Test einzusetzen, mag aus folgenden Überlegungen heraus plausibel sein: Aufgrund der beschränkten Speicherkapazität stand in dem Graphem-Phonem-Konversionsprogramm das 'h' wohl im absoluten Anlaut, nicht aber im wortmedialen Silbenanlaut (z.B. a'ha) zur Verfügung. Dieselbe Begründung gilt für 'pf', das nicht als eigens kodierte Affrikata (wie 'tsch' und 'z' = 'ts') verfügbar war, sondern als Konsonantencluster 'p''f' kombiniert worden wäre (was andererseits ja auch für 'tsch' und 'ts' möglich scheint). Bis zur Klärung dieser Fragen (etwa Aufnahme von Wörtern mit medialem 'h' bzw. mit initialem 'ch' in ein Ausnahmewörterbuch, oder Rekombination aller Affrikaten aus ihren Bestandteilen) halten wir den Verzicht auf die fraglichen Laute in einem ersten systematischen Evaluationstest, wie ihn unsere Untersuchung darstellt, für vertretbar.

Als vokalische Umgebung dienten für jeden Konsonanten die fünf Vokale i,e,a,o,u; bei den VCV-Stimuli waren jeweils der dem Konsonanten vorangehende und der folgende Vokal gleich. Alle Stimuli wurden mit schwebender Intonation (d.h. ohne Grundfrequenzänderung) generiert und auf Band aufgenommen. Das Testmaterial bestand also für die CV-Serie aus 17 (C) x 5 (V) = 85 Stimuli, für die VCV-Serie aus 19 x 5 = 95 Stimuli, die jeweils innerhalb jeder Serie so randomisiert wurden, daß nie zwei Stimuli mit gleichem C und/oder gleichen V aufeinander folgten. Das Intervall zwischen den Stimuli betrug 3 sec.

Das Testband mit beiden Serien (zuerst CV, dann VCV) wurden in einer Sitzung 20 Versuchspersonen über Raumlautsprecher dargeboten; die Versuchspersonen (österreichische Studenten; Durchschnittsalter ca. 21 Jahre; 15 weiblich, 5 männlich) wurden darüber informiert, daß es sich um synthetisierte Laute handelt und sollten auf einem Antwortbogen in deutscher Orthographie notieren, was sie zu hören vermeinten (free choice). Dabei war auch eine Null-Lösung (= nicht erkannt) zugelassen, um wirklich 'free choice' zu bieten. Die Eintragungen auf den Antwortbögen wurden dann mittels Personal-Computer in einem eigens zusammengestellten Ablaufschema bestehend aus Such-, Sortier-, Zähl- und Statistikprogramm ausgewertet.

Ergebnisse

Die Ergebnisse dieses Evaluationstests lassen sich nach mehreren Gesichtspunkten ordnen: 1) Verständlichkeit der Konsonanten insgesamt ausgedrückt als Anteil aller richtigen Urteile an der Gesamtzahl der abgegebenen Urteile; 2) Verständlichkeit der einzelnen Konsonanten ausgedrückt durch die jeweilige Anzahl der richtigen Urteile; 3) Verwechslungshäufigkeit der einzelnen Konsonanten ausgedrückt durch die jeweilige Art und Anzahl der Fehlurteile. Alle drei Beurteilungsaspekte kommen sowohl für den Gesamttest als auch ge-

trennt nach der jeweiligen (initialen und medialen) Position des Konsonanten in Betracht. Eine weitere Differenzierung läßt sich noch durch die Variation des Umgebungsvokals gewinnen.

Eine ausführliche und anschauliche Darstellung der Untersuchungsergebnisse nach allen diesen Aspekten mit Tabellen und/oder Grafiken würde den Rahmen dieses kurzen Berichtes sprengen. Es werden daher (in Tabelle 1 und 2) nur die Verwechslungsmatrizen mit den Rohdaten für jede der beiden Konsonantenpositionen vorgelegt, aus denen sich (abgesehen von der hier vorläufig nicht weiter verfolgten Differenzierung durch die vokalische Umgebung) alle übrigen Resultate ergeben (und zum eventuellen Nachvollzug errechnen lassen). Zur Notierung der Konsonanten in diesen Matrizen und in den folgenden Ausführungen noch eine Anmerkung: Da für die Beschriftung der Matrix nur jeweils zwei Zeichenpositionen je Kolonne zur Verfügung standen, werden für 'sch' und 'tsch' die Zeichen s' und ts' verwendet; das Graphem 'z' wird durch ts ersetzt, um die charakteristische Nähe der beiden Affrikaten ts und ts' hervorzuheben.

In den Verwechslungsmatrizen sind von oben nach unten die zur Beurteilung vorgegebenen Laute und von links nach rechts die Urteilsvarianten aufgelistet. Die letzte Kolonne enthält die Null-Lösungen. Die erste Datenzeile von Tabelle 2 beispielsweise besagt also, daß 'p' von den Versuchspersonen 32mal als p, 13mal als t, 4mal als k, ..., 3mal als unverständlich (Null) beurteilt wurde. Die Quersumme jeder Zeile beträgt 100 (errechnet aus 5 vokalischen Umgebungen mal 20 Urteilen je Konsonant).

Aus diesen Verwechslungsmatrizen lassen sich folgende Ergebnisse ableiten: 1) Die Verständlichkeit aller Konsonanten insgesamt beträgt 35%, wobei sich kein positionsabhängiger Unterschied zeigt. 2) Die Verständlichkeit der einzelnen Konsonanten für beide Positionen zusammen weist folgende Reihenfolge auf: 60-50% /j,d,s',m,l,s/; 40-30% /w,t,ts/; 30-20% /r,p,k,b,ts',f/; unter 20% /g,n/. Für die beiden nur inlautend getesteten Konsonanten, nämlich 'ch' und 'ng', wurden 44% bzw. 0% verzeichnet. Bei der Verständlichkeit der einzelnen Konsonanten zeigt sich in einigen Fällen ein deutlicher Unterschied zwischen den beiden Positionen. In VCV-Stimuli sind die Laute /k,b,d, s,s',ts'/ zwischen 15% und 30% besser verständlich als in den entsprechenden CV-Stimuli; in CV-Stimuli hingegen sind es die Laute /m,j,l,ts/, die zwischen 20% und 40% besser beurteilt werden als in entsprechenden VCV-Stimuli.

Alle diese Angaben zur Verständlichkeit bzw. Identifikationsquote der Konsonanten sind an sich schon recht aufschlußreich (zumal sie in Einklang stehen mit den durch ungarische TTS-Systeme erzielten Ergebnissen für ein- und zweisilbige Logatome; vgl. GOSY/OLASZY, 1983 bzw. OLASZY/GORDOS, 1988). Sollen aber die Ergebnisse eines solchen Evaluationstests nicht bloß den erreichten Standard eines TTS-Systems demonstrieren, sondern vielmehr durch diagnostische Analyse die Grundlage zu verbessernden Modifikationen des Systems liefern, ist mit der Angabe von Identifikationsquoten nur wenig gedient.

Ein weitaus nützlicherer Beitrag ist eher durch jene Information zu erwarten, die sich aus den Verwechslungsmatrizen ergeben. Aus der Art und Anzahl der jeweiligen Fehlurteile über einen Stimulus läßt sich nämlich besser auf die für notwendige Modifikationen relevanten Parameter schließen.

Die Verwechslungshäufigkeiten für die jeweilige Konsonantenposition lassen sich aus Tabelle 1 und 2 entnehmen. Für die folgenden Angaben wurden die Ergebnisse aus beiden Tabellen zusammengezogen und insofern vereinfacht, als nur die jeweils fünf häufigsten Urteilsvarianten aufgelistet werden. Dadurch lassen sich allzu große Urteilsstreuungen und eventuelle Ausreißer eliminieren. Im einzelnen waren die nachstehenden Verwechslungscluster zu beobachten (der Stimulus steht als Großbuchstabe gefolgt von den fünf Urteilsvarianten):

P - /b p t Ø d/ T - /t p Ø d g/ K - /k p g Ø b/ B - /b f g Ø h/
D - /d b m l n/ G - /d Ø g l b/ M - /m b w Ø n/ N - /m b w l n/
NG - /m Ø d b n/ F - /ch f h p b/ S - /s s' Ø j ch/ S' - /s' f ch Ø p/
CH - /ch s' h f Ø/ J - /j Ø l w s'/ W - /w m n d Ø/ R - /r b Ø w d/
L - /l r Ø w b/ TS - /ts ts' s t Ø/ TS' - /ts' p k t pf/

Aus diesen Gruppierungen ergeben sich zwei Hauptziele für die weitere Entwicklung des TTS-Synthesesystems FLEX-DEUTSCH: 1) Modifikation der relevanten akustischen Parameter zur Optimierung der Diskriminierbarkeit innerhalb von nach Artikulationsmodus und Stimmtonbeteiligung definierten artikulatorischen Klassen. 2) Auditive Überprüfung der optimierten synthetischen Laute durch forced-choice-tests innerhalb der oben genannten Klassen. Der nächste Schritt nach allenfalls weiteren verbessernden Modifikationen wäre dann die systematische Verständlichkeitsprüfung an Hand von deutschen Minimalpaaren.

Literatur

1. GOSY, M.-OLASZY, G.: The perception of machine voice. Nyelvtudományi kozlemények. Budapest, 1983.
2. OLASZY, G.: Die Anwendungen des Flex-Deutsch Sprachsynthesesystems in phonetischen Forschungen. Hungarian Papers in Phonetics 19. 1988, 34-46.
3. OLASZY, G.-GORDOS, G.: On the speaking module of an automatic reading machine. Hungarian Papers in Phonetics 17. 1987, 163-191.
4. POLS, L.C.W.: A joint Dutch research program for developing a high-quality text-to-speech synthesis system. Proc. of the Inst. of Phonetic Sciences Amsterdam 12. 1988a, 9-17.
5. POLS, L.C.W.: Improving synthetic speech quality by systematic evaluation. Proc. of the Inst. of Phonetic Sciences Amsterdam 12. 1988b, 19-27.
6. TERKEN, J.-LEMEER, G.: Effects of segmental quality and intonation on quality judgements for texts and utterances. Journal of Phonetics 16. 1988, 453-457.

cv	p	t	k	b	d	g	m	n	ng	f	s	s'	ch	j	w	r	l	ts	ts'	pf	h	Ø
p	24	9	1	25	7	5	3	3		4					1						1	17
t	18	39	8	2	2	9	1			3						1					3	14
k	18	5	18	6	1	26	1			2					1			1	1		2	18
b	15	5	1	12		1	3	1		33	2				4	1					14	8
d				23	46		10	9						1			7					4
g	3			9	15	9	8	9		5				1	2	1	17		3			18
m	1			4	1		72	8							8		2					4
n				2			65	9			1				10	1	8					4
ng																						
f	10	1	3		1	1				29		2	11		1		1			8	24	8
s		3				1					43	9	2	16		5	2	1	1			17
s'	9	2	1							17	1	44	4		1	2		1	6		3	9
ch																						
j			1	2	5	5								71	1		7			1		7
w				1	2		26	16						4	40	1	3	1				6
r	7	6		10	4	1	3						1		9	31	9			1		18
l	1			10	4			1			1			2		7	65	1				8
ts		3	4				2				18							45	23	1		4
ts'	24	26	25	1						1					1			1	9	5		7

Tab.1: CV-Stimuli (Verwechslungsmatrix)

vcv	p	t	k	b	d	g	m	n	ng	f	s	s'	ch	j	w	r	l	ts	ts'	pf	h	Ø
p	32	13	4	34	8	6																3
t	18	31	4	10	14	4				5				2					1	1		10
k	19	4	38	11	2	8				4			1							1	3	9
b				44	14	18	1						1		6	3	1				2	10
d	1			22	68	4											3					2
g			1	9	23	18	1			6	1	6	13	2	2	3	2				2	11
m		2		26	3	8	32	1						1	11	4	3				1	8
n	1			30	13	4	30	4						1	5	2	6				1	3
ng	4			9	11	5	27	7		1		1	2	1	3	4	5				1	19
f	5			11						17	10	2	40								13	2
s											58	35	7									
s'										23		62	13								1	1
ch										3		33	44								15	3
j					2	3	3	1			8	9	1	51	9	1	6					6
w		2		4	11	8	6			1				2	39	9	7				5	6
r				21	10	6	2							3	11	30	5				2	10
l				2	5	2	1							4	16	20	37				1	12
ts	7	1	1	3	3	4				1	3	4	1		1			25	42	1		4
ts'	23	3	7	3		1				3						1			44	10		5

Tab.2: VCV-Stimuli (Verwechslungsmatrix)

INTERAKTIVE SPRACHERKENNUNG

Reinhold GREISBACH
Institut für Phonetik
Universität zu Köln, Köln, Deutschland

Spracherkennung, d.h. die Transformation eines Sprachsignals in eine Folge von (niedergeschriebenen) Zeichen, ist eine der ureigensten Aufgaben des Phonetikers. Zugleich ist Spracherkennung z.Z. ein Forschungsschwerpunkt im Rahmen der Mensch-Maschine-Kommunikation. Automatische Spracherkennungssysteme (ASE-Systeme) sind bereits heute in der Lage Wörter aus einem recht großen Wortschatz, selbst wenn sie von verschiedenen Sprechern geäußert wurden, nahezu in Echtzeit und fast fehlerfrei zu erkennen.

Jedoch sind diese Fähigkeiten der ASE-Systeme zur Lösung der obengenannten Aufgabe des Phonetikers nicht unbedingt hilfreich. Seine Aufgabe, genannt phonetische (impressionistische, enge) Transkription, erfordert eine (auditive) Spracherkennung, die zum Teil jenseits der von den heutigen ASE-Systemen erreichten Grenzen liegt, nämlich eine möglichst detaillierte Lauterkennung bei einem unbegrenzten Sprach- und Wortinventar. An einen Ersatz des Menschen bei der phonetischen Transkription durch ein ASE-System ist deshalb z.Z. noch nicht zu denken.

Dagegen lassen sich die Funktionen der Hilfsgeräte, die der Phonetiker oder allgemeiner der Transkribent für die phonetische Transkription benötigt, bereits auf dem Computer simulieren. Ein solches Programm zur computer-unterstützten Transkription (cuT) integriert die Simulation eines (Bandschleifen-)Tonbandgerätes zur akustischen Wiedergabe des zu transkribierenden Signals (D/A-Wandlung eines vorher digitalisierten und abgespeicherten Sprachsignals auf Knopfdruck) sowie den Vorgang des Niederschreibens (Notation mit Hilfe eines Textverarbeitungssystems mit phonetischem Zeichensatz und Ausgabe auf Drucker). Die Benutzung eines solchen cuT-Systems anstelle der üblichen transkriptorischen Hilfsgeräte sollte zu einer Reduktion der Arbeitszeit (Kopieren, Schneiden und Kleben von Tonbandschleifen entfällt) und möglicher Übertragungsfehler beim Kopieren der handschriftlichen Notation (z.B. mit einer Kugelkopfschreibmaschine) führen.

Darüber hinaus bietet ein cuT-System dem Anwender neue, im Rahmen der Transkription bisher kaum genutzte Hilfsmittel, wie die Segmentatorfunktion (nur ein ausgewählter Signalausschnitt zwischen zwei Markierungen wird hörbar), ein akustisches Referenzlautinventar (als auditive Vergleichsmuster einsetzbar) oder den Rückgriff auf akustische Informationen (graphische Darstellung des Sprachsignals als Oszillo-

gramm oder Spektrogramm). Dazu kommen neue Möglichkeiten, bedingt durch die Integration aller Funktionen in einem Programm. So können z.B. - nachdem der Transkriptionstext mit dem Textverarbeitungssystem erstellt ist und die zu den Symbolen gehörigen jeweiligen Segmentgrenzen im Signal markiert sind - die Symbole bzw. Symbolfolgen direkt aus dem Text "abgehört" werden. Dadurch läßt sich u.a. überprüfen, ob ein Segment überhört oder eines zuviel notiert wurde.

Für eine detaillierte Beschreibung der Anforderungen an ein solches cuT-System siehe (1). Dieser Anforderungskatalog diente als Grundlage für die Entwicklung des cuT-Systems NOTAT (erstmals 1987 für einen Computer der PC-Klasse vorgestellt (2)). Die Benutzer von NOTAT bestätigten den erwarteten Zeitgewinn beim Transkriptionsvorgang, falls nur die transkriptorischen Grundfunktionen (Abhören und Notieren) benutzt wurden. Ein intensiver Einsatz der neuen Zusatzfunktionen führte jedoch zu einer starken Erhöhung der Arbeitszeit.

Hier sollte nun die Kombination eines ASE-Systems mit einem cuT-System wieder zu einer Reduktion des Zeitaufwandes führen. Das ASE-System (genauer der akustisch-phonetische Modul eines ASE-Systems) arbeitet dabei als automatischer Präprozessor für den manuellen cuT-Vorgang.

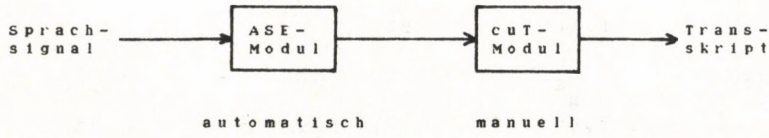


Abb. 1: Kombiniertes ASE-/cuT-System

In einem kombinierten ASE-/cuT-System wird das digitalisierte Sprachsignal zunächst im ASE-Moduls automatisch analysiert, segmentiert und mit Symbolen versehen. Der so gewonnene Text kann danach mit dem cuT-Modul des Systems auditiv überprüft und ggf. manuell verbessert werden.

Für die Konzeption des ASE-Moduls ergeben sich aufgrund der besonderen Anforderungen in einem solchen System folgende Vereinfachungen:

Die Signalverarbeitung muß nicht in (Quasi-)Echtzeit erfolgen. Der automatische Erkennungsprozeß kann hierbei in der Vorbereitungsphase vor dem eigentlichen Transkriptionsprozeß ablaufen. Dies gilt insbes. für die zeitaufwendigen Teile der Erkennung, wie Signaltransformationen (z.B. digitale Filterung, Fourier-Transformation) und Parameterextraktionen (z.B. LPC-Analyse, Formantbestimmung). Zudem können auch verschiedene konkurrierende Verfahren zur Bestimmung eines einzigen Parameters ablaufen.

Der automatische Erkennungsprozeß muß nicht sprecherunabhängig sein. Bei der Transkription werden meist längere Äußerungen eines einzigen Sprechers bearbeitet, so daß das Erkennungssystem zu Beginn darauf abgestimmt werden kann. Die akustischen Vergleichsparameter für die Klassifikationsprozeduren können z.B. durch Analyse der Signale aus dem Referenzlautinventar des cuT-Moduls gewonnen werden, falls sie vom gleichen Sprecher stammen.

Hinsichtlich der Genauigkeit des automatischen Prozesses sind an den Segmentationsprozeß größere Ansprüche zu stellen als an den Klassifikationsprozeß. Denn Symbole als die Ergebnisse des Klassifikationsprozesses lassen sich recht leicht mit der Textverarbeitungseinheit des cuT-Moduls verändern.

Die Verwendung eines an das cuT-System gekoppelten ASE-System bedeutet also nicht notwendig die Abkehr vom auditiven Charakter der Transkription. In welchem Maße sich der Transkribent auf die akustischen Analysedaten und damit auf die vom automatischen Prozeß vorgeschlagene Symbolfolge verläßt, bleibt allein ihm überlassen.

Besteht in einem kombinierten ASE-/cuT-System die Möglichkeit auch den Erkennungsprozeß manuell zu beeinflußen, so ergibt sich ein interaktives Spracherkennungssystem (ISE-System). Dabei sind ggf. die zeitaufwendigen Teile des Erkennunsprozesses abzutrennen und weiterhin in einer Vorverarbeitungsphase durchzuführen.

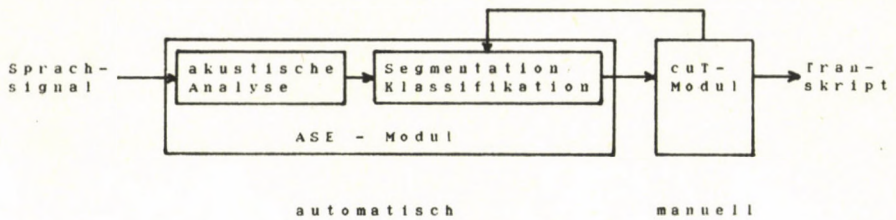


Abb. 2: ISE-System

Ein ISE-System erlaubt dem Benutzer manuell auf die Daten der akustischen Analyse und auf Teile des automatischen Erkennungsprozesses selbst einzuwirken. Dabei ist insbes. an eine Beeinflussung des Segmentations- und Klassifikationsprozesses zu denken. Der Benutzer kann hierbei nach einer Änderung (z.B. von Schwellwerten beim Segmentationsprozeß) diese Teile des automatischen Erkennungsprozesses unter den geänderten Voraussetzungen nochmals ablaufen lassen. Aber auch durch Manipulation der automatisch gewonnenen akustischen Parameter (z.B. falsch erkannter Formanten) kann der Benutzer eines ISE-Systems den Erkennungsprozeß beeinflussen.

Zudem können die akustischen Parameter (nachdem sie überprüft sind) zusammen mit den transkribierten Symbolen archiviert werden und für spätere statistische Auswertungen, Resyntheseverfahren oder Verbesserungen des ASE-Moduls benutzt werden, so daß das ISE-System als Erfassungssystem für eine phonetische Datenbank genutzt werden kann.

Mit Programm INTER-NOTAT soll die prinzipielle Funktionsfähigkeit des ISE-Konzeptes geprüft werden. Die benötigte Hardware besteht aus einem Computer der PC-Klasse (ATARI ST), einer A/D-D/A-Einheit (Eigenentwicklung IPKöln) und einem Matrixdrucker (NEC P6). INTER-NOTAT basiert auf dem cuT-System NOTAT (vgl. auch (3)), ergänzt um eine Manipulationseinheit für den Segmentations- und Klassifikationsprozeß und eine Funktion zur Bearbeitung der akustischen Analysedaten. Als ASE-Modul für INTER-NOTAT ist zunächt ein einfaches System implementiert, wobei insbes. auf die Effektivität der Segmentation geachtet wurde.

Referenzen

1. GREISBACH, R.: Computers and the transcription of speech. In: Jung, U.O.H. (Hrg.): Computers in applied linguistics and language teaching. Frankfurt a.M. 1988. S. 147-153.

2. GREISBACH, R.: Computerunterstützte Transkription. In: Spillner, B. (Hrg.): Angewandte Linguistik und Computer, Kongreßbeiträge der 18.Jahrestagung der Ges. f. Angew. Ling. 1987, S. 88-89.

3. NOTAT-Handbuch, Institut für Phonetik, Universität Köln, 1989

AN APPROACH TO ARTICULATORY REPRESENTATION OF SPEECH SIGNAL ON THE BASIS OF ITS APPROXIMATE PARAMETRIC ANALYSIS

Ryszard GUBRYNOWICZ

Laboratory of Speech Acoustics

Institute of Fundamental Technological Research

Polish Academy of Sciences

Warsaw, Poland

1. Introduction.

Many speech researcher are tempted to use linguistic category names (phonemes, vowels, consonants and others) to observations and measurements made on selected parts of the acoustic speech signal. This type of misuse is particularly frequent when they are trying to formulate recognition rules of linguistic units based on parameterical representation of the speech signal. In fact, linguistic categories are abstract by nature and have no physical meaning. B. REPP [1] warned against mixing of terms from different levels, especially, against applying to the analysed parts of the speech waveform the abstract notions such as phonemes. Linguistic units are important concepts for describing and explaining the structure of linguistic items, but it is impossible to represent them on the acoustical level. The closest level to the latter is the articulatory one which is intermediary between the acoustic representation and linguistic description of speech. However,the relations between articulators' configuration and acoustic characteristics of the speech signal are often ambiguous and in many cases they could be evaluated in approximate way only. For these reasons we have chosen fuzzy rules to describe in broad categories the basic articulatory classes represented in the universe of acoustic cues [2,3].This paper presents a system of speech recognition with an extended speech articulatory representation, which is under development. The word was adopted as a unit of recognition and is described in form of a sequence of articulatory units. A vocabulary of 60 words was chosen to control by speech the transmission of data to/from computer and their processing.

2. Overview of the system of approximate articulatory description.

A general scheme of articulatory representation of Polish speech sounds was elaborated. Each sound was defined in the articulatory space with the main variables as follows:

a) the manner of articulation,
b) the place of articulation,
c) the height of articulation.

2.1 Broad description of the manner of articulation

Until now a two - level articulatory description scheme was used (Fig.1). At the first level, a small set of primary articulatory classes was defined on the basis of broad evaluation in acoustic space. Four different manners of articulation were distinguished,

namely, resonants (RS), strideents (SS), unvoiced closures (UC) including also silences and the more general class called obstruent

SCHEME OF THE TWO-LEVEL ARTICULATION DESCRIPTION OF THE SPEECH SIGNAL

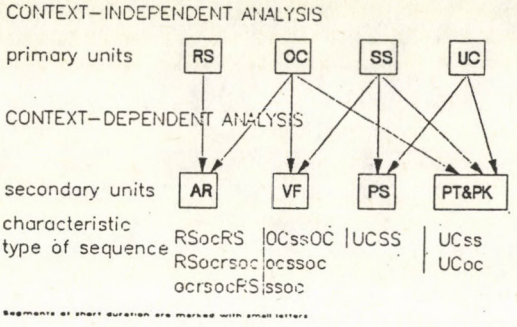


Fig. 1

candidates (OC). The distinction between these speech segment categories has a very clear articulatory meaning and could be accomplished easily in acoustic space.

For manner description, three of measured parameters were used, namely, the amplitude envelope level and levels in specified low-pass (LP) and high-pass (HP) frequency bands, and 4 supplementary parameters were calculated. The first stage of manner analysis was context - independent and based on the examination of several functions defined over measured and calculated parameters. They were characterized as "low", "mean" or "high" by imposing on their values fuzzy restrictions used for formal definitions of the above articulatory classes. In the proposed system of mapping acoustic cues into articulatory features the values of these functions were characterized with taking into account also natural ambiguities

AO = amplitude envelope lev.

LP = low-passed ampl. lev.

HP = high-passed ampl. lev.

$DA = AO_{max} - AO$

$DL = LP_{max} - LP$

$HL = HP - LP$

$HA = HP - AO$

$RS = l(DA) \circ [l(DL) \vee l(HL)]$

$SS = h(HL) \wedge h(HA)$

$UC = l(AO) \wedge l(LP) \wedge l(HP)$

$OC = \overline{UC} \wedge \overline{SS} \wedge \{[m(DA) \vee h(DA)] \circ [m(DL) \vee h(DL)]\}$

where $\circ$ strong conjunction $\quad a \circ b = \max(a+b-1, 0)$

$\wedge$ rigid conjunction $\quad a \wedge b = \min(a, b)$

$\vee$ disjunction $\quad a \vee b = \max(a, b)$

Fig. 2

l(), m(), h() - "low", "medium" and "high" functions.

existing in speech signal. This was another reason to use linguistic

description of parameters' variations instead of describing them very precisely. Fig. 2 presents formalized naming relations defining four basic articulatory classes independently of their context.

The second part of manner analysis which is context - dependent, is devoted to detection of some characteristic sequences of labels and to recode them into new articulatory categories or to correct errors in the primary description. For some specific cases a return to more detailed parametric analysis was also possible. At this stage of analysis the position of chosen sequences in the word, their articulatory environment and durations were taken into account. The final result of manner analysis was a string of labels obtained for the input word (see the fig.3).

2.2 Broad description of height and place description

The approximate description of height and place of articulation was based on formant frequencies analysis. For height of articulation three fuzzy functions defined over first formant frequency as "low", "medium" and "high" were used. The place of articulation was described after second formant evolutions as "front" or "back". The articulatory description was carried out for each 10 ms frame containing voiced part of the speech signal, e.g., for segments described as **RS, OC, AR** and **VF**. An example of height and place description results are given on the fig. 3.

This description was used for more detailed articulatory representation of speech signal and farther sub-classification of above mentioned broad articulatory classes, especially, resonants. At this stage of analysis a verification of primary description could be also possible. It is important especially in the case when a short segment of the class **OC** is at the beginning of the word. The fig. 3 presents the case of a non detected earlier back plosive /k/ spoken at the beginning of the word. The supplementary analysis of height and place analysis support the hypothesis that the local maxima in low-pass and envelope amplitudes at the beginning of the word are significant. For other positions, a sequence **UCOC** with short duration of the second segment (about 30-40 ms) is typical for unvoiced back plosive and the analysis of above parameters is unnecessary.

3. Generation of reference sequences of the lexicon

An important part of the system is the package of procedures for automatic generation of broad articulatory description of words included to the lexicon. Each word introduced from the keyboard is automatically translated by means of context dependent rules, from its orthographic form into a number of articulatory transcriptions. The knowledge of the articulatory rules which takes into account also the main speaker variations permits to correct in some mesure the generated word patterns used as reference sequences in process of recognition. Moreover, for a given vocabulary, an analysis of confusability of word patterns could be done automatically. On the fig. 4 an example of reference lattice of articulatory segments obtained for the polish word "monitor" is presented. At the bottom of the figure, for the last reference sequence **RSUORS** the results of the analysis of words confusion for the 60 words lexicon are given.

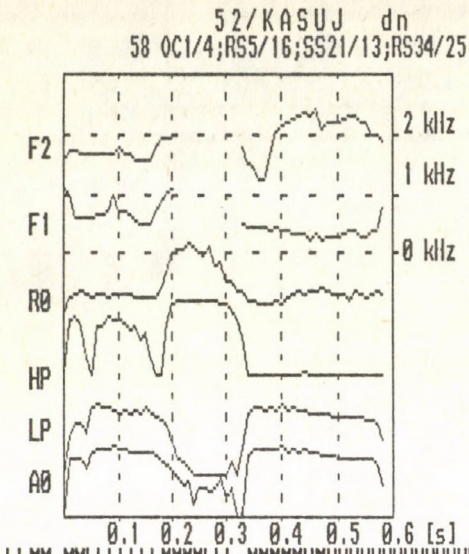
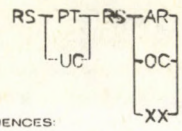


Fig. 3

THE ARTICULATORY CHART OF THE WORD

typed word: monitor

Lattice of articulatory segments



REFERENCE SEQUENCES:
1. RS PT RS AR
2. RS UC RS AR
3. RS PT RS OC
4. RS UC RS OC
5. RS PT RS
6. RS UC RS

RESULTS OF CONFUSION ANALYSIS
FOR REFERENCE SEQUENCE: RS UC RS

WORD CANDIDATES

14. PULPIT /pulpit/ - (control desk)
15. MONITOR /monitor/ - (monitor)
55. KONTYNUUJ /kontynuuj/ - (continue)

Fig. 4

4. Results

The average error of articulatory description of 60 words spoken in isolation by 13 speakers (two of them with neglectful articulation) was about 10 %. However, for 9 speakers this score was about 6%. In general, the errors (insertion, deletion and substition) in detection of articulatory units varied from 0 to 25% (for bad speakers). The mean scores were for RS < 1%, UC - 1.4%, OC - about 26%, SS - < 0.5%, AR - 24%, VF - 8% and for plosives - 12%. The recognition error for a set of 30 words was about 5%.

The presented system of isolated word recognition is very simple and flexible. One of the main advantage of the presented approach is the possibility to use the phonetic and phonological knowledge on the articulatory level of speech signal description. Applying on this level the linguistic information, the rules of recognition could be largely verified without processing the real speech signal, as it was presented above in the example of word confusion analysis.

References.

[1] B. REPP, On levels of description in speech research, Haskins Lab. St. Rep on Speecz Res., SR-65 (1981), 217-222.

[2] R. GUBRYNOWICZ, Mapping acoustic cues into phonetic features, Proc. IV FASE Symp. on Acoustics and Speech, Venezia,1981, ESA-Roma,301 - 304.

[3] W. WIĘZLAK, R. GUBRYNOWICZ, Articulatory description of speech signal in isolated word recognizer, Proc. of IEEE ICASSP, Paris, 1982,529-534.

TEACHING OF HUNGARIAN TO FOREIGNERS BY SPEAKING COMPUTER

KOUTNY Ilona & OLASZY Gábor
Linguistic Dept. of ELTE Inst. of Linguistics, HAS
Budapest, Hungary

1. Introduction

In the domain of CALL (Computer Aided Language Learning) the main role of the computer concentrates on *exercising grammar and vocabulary* - this provides the basis for speech capability. But nowadays the aim of language instruction is to form communicative competence in the pupils. How can a silent teaching aid teach speech?

Automatic speech synthesis, available already for several languages, could contribute to the development of auditory capability. Though the speech quality cannot rival that of a tape recorder, integrated into the *computerised learning process* (task presentation, assistance in the solution, checking and evaluation of the solution and individualised continuation), synthetised speech offers a better means of coping with speech.

The dictation programs check spelling and oral comprehension. Other tasks involve grammatical structures as well. The Hungarian language is a challenge both for pupil and computer because of its agglutinative nature. The kernel of these programs is a noun-form generator. The noun-forms are exercised in simple sentences using the methods: completion, transformation and questions.

The dictionary and the texts used for the programs can be extended and changed by the teacher, so they are so-called authoring programs. They can be applied at different levels with an appropriate text and dictionary.

2. Computer generated Hungarian speech

Synthetised speech is very rarely used in CALL (one of the first ones was Sherwood, 1981 [10]). Looking for the motives for this we have to know a little about the mechanism of automatic speech generation. (1) Limited vocabulary speech synthesis uses encoded speech, which requires a big storage capacity and the utterances have to be known and stored in advance. (2) Speech synthesis with unlimited vocabulary, the so-called *full text-to-speech synthesis* which can transform any text into speech by rules.

For teaching purposes (1) is not convenient because a lot of words are to be taught and exercised in different contexts, and also the storage capacity on microcomputers is limited. (2) can satisfy the requirements of language teaching. The price to pay for the capability to generate every thing is a loss in the naturalness of the speech. The generated speech can be used in learning only if its quality is good enough.

The rule based speech synthesis is carried out in three main steps: (1) conversion of graphemes into phonemes; (2) parameter determination of speech units; (3) superposing of prosody. So the complexity of generation and the speech quality depends on (1) the regularity of the orthography, (2) the complexity of the sound system and (3) the regularity of accent of the given language.

In this respect Hungarian is in a relatively good position to be well synthesisable. The Hungarian orthography is quite regular and the accent is always on the first syllable. The speech sounds and sound transitions can be described by about 1000 rules [8]. A good quality of synthetised Hungarian has been reached and it is supported by intelligibility tests [2].

A rule based text-to-speech synthesis system for Hungarian (SCRIPTOVOX) and for Esperanto (ESPAROL) developed through the cooperation between the Institute of Linguistics and the Technical University of Budapest (1983-88) provides the speech for our programs. A small device plugged into a microcomputer Commodore 64 or 128 generates the mentioned languages and it can be activated from Basic. The IBM PC version comprises a plug-in board and it is programmable in Pascal.

3. Role of synthetised speech in CALL

Communication with computer is more friendly if it is done by voice. Learning foreign languages the speech becomes crucial. The communicative teaching emphasises the development of *communicative abilities* (receptive ones: reading and listening, and productive ones: writing and speaking). In order to teach speech we have to apply speech. In general the teaching programs can develop only two capabilities: the writing and the reading. First of all they can exercise the grammar and the vocabulary which are the basis for accuracy in speech [3]. The computer cannot yet cope with the whole of speech communication, but by making use of the synthetised speech it can contribute to the development of *auditory ability*.

The role of computer speech in the learning process is described by Koutny in [6]. The task is communicated by voice and the pupil has to write his answer. The evaluation is by voice too (e.g. *O.K. You've done it well! Try it again! I am sorry, you failed.*). Often the good solution is confirmed by its pronunciation. Tasks can be listened to once more.

The speech can assume a *primary role* in some CALL-tasks (dictation, telling stories, asking questions, definitions by voice etc) or a *supplementary role* when the task appears on the screen (texts to be completed, transformed, built etc) and after the pupil's answer the good solution and only the good one is pronounced. Usually the computer communicates with the pupil by voice.

The first teaching program package according to these considerations was elaborated and tested in a beginners' course for Esperanto [4,5,7]. Now teaching programs of Hungarian for foreigners are developed. Tests are planned during the Summer University Courses in Debrecen with the collaboration of KLTE.

4. Speaking programs

The most evident task for a speaking computer is the *dictation*. The system chooses sentences randomly from a set of sentences one after the other and dictates them. The pupil can listen to them again if he wants to, and has to write them. The spelling of the pupil is then checked. If something is incorrect the system displays the correct part up to the mistake and requests a correct continuation repeating the whole sentence by voice. In a modifying mode the sentences can be listed, modified and extended.

A *speak and spell* like program for Hungarian is also available [1]. It exercises words. We have to remark that the sentence level intelligibility of the system is comparable to that of human speech, in the case of isolated words it is a little bit less [2].

The *comprehension* exercise involves a small story. The pupil has to listen to the computer and afterwards to give one-word answers to some oral questions relating the story. The questions can be repeated. After a good answer the computer speaks the whole relevant sentence from the story. If the answer is incorrect, the question is repeated once more and the pupil can try it again. After the second incorrect answer the sentence containing the answer is given on the screen. The teacher can prepare stories with questions and answers. So the program can offer several stories one after the other.

The *Anagram* game exercises vocabulary. It says the definition of a word and displays its letters. If the pupil's answer is not correct, it says the definition once more and displays the next letter of the answer. The word and the definition set can be easily changed.

5. Exercising Hungarian noun-forms

Hungarian is difficult to learn for foreigners not familiar with agglutinative languages because the complexity of noun and verbal endings modulated by the vowel harmony poses a big problem. Hungarian uses suffixes instead of many prepositions and possessive determiners. The *noun-form generator* can generate the following forms e.g. from the word *ház* [house]:

házat (Acc.), *házból* [from a house], *házban* [in a house], *házba* [into a house], *házról* [down from a house], *házon* [on a house], *házra* [onto a house], *háztól* [away from a house], *háznál* [at a house], *házhoz* [towards a house], *háznak* [to a house], *házig* [up to a house], *házért* [for a house], *házzal* [with a house], *házzá* (Factive)

and similarly their plural forms. The stems can be classified depending on their behaviour before endings. We use a simplified version of the Papp grouping [9]. We distinguish 5 stem types (others are handled as exceptions). The affixation type relates on the ending in Accusative, Plural and Possessive.

The nouns are stored in the noun dictionary with their grammatical information. During the preparing of the tasks

nouns are put into the dictionary, the system helps to define their type. The whole noun paradigm can be listed. It is better to exercise noun-forms in sentences, therefore we build simple sentences with them using only a few verbs in Singular 3Prs and contained in a small verb dictionary. This also knows which complements can be associated with each verb. This kernel serves for several exercises:

Answer! A question appears on the screen with the noun to be used in the one-word answer, e.g.

Hová megy Péter? (iskola)

Where is Peter going to? (school)

After the correct answer the whole sentence is uttered:

Péter iskoldba megy.

Ask! A simple sentence appears on the screen. The pupil has to make a question about the highlighted word, e.g.

A fiú könyvet olvas. [The boy is reading a book.]

After a good answer the system utters the whole question:

Mit olvas a fiú? [What is boy reading?]

Transform! The sentence on the screen must be transformed into the Plural, e.g.

A lány sétál az utcán. [The girl walks along the street.]

The good answer is spoken: *A lány sétál az utcákon.*

These pilot programs integrate the synthetised speech into CALL. Further researches aim to involve other structures of Hungarian to be taught by speaking computer.

References

[1] Gordos Géza (1987): *Helyesírási oktatóprogram személyi számítógépre* [Spelling teaching program for personal computers]. In: AV kommunikáció 87/3-4 pp 150-151

[2] Gósy Mária, Olaszy G. (1983): *A gépi beszéd megértése* [Intelligibility of computer speech]. In: Nyelvtud. Közl. 85/1 pp 93-104

[3] Kecskés István (1986): *Complex, Cyclical, Generative Programs to Teach Grammar.* In: Linguistics and Methodology in CALL. Ed. I. Kecskés & F. Papp. Budapest: SZÁMALK pp 37-52

[4] Kisfaludy Katalin (1988a): *Apliko de sintezita parolo en la komputila instruado de Esperanto.* Thesis

[5] Kisfaludy K. (1988b): *Instruado de Esperanto per parolanta komputilo.* In: Internacia Pedagogia Revuo 1988/4 pp 7-9

[6] Koutny Ilona (1988): *Számítógépes beszédelőállítás alkalmazása a nyelvoktatásban* [Application of computer speech generation in language teaching]. In: AV kommunikáció 89/2

[7] Koutny I., Olaszy G., Kisfaludy K. (1988): *Esperanto speech synthesis and its application in language learning.* In: Hungarian Phonetic Papers N 19 pp 47-54

[8] Olaszy Gábor (1989): Elektronikus beszédelőállítás [Electronic speech generation]. Budapest: Műszaki Kiadó

[9] Papp Ferenc (1975): *A magyar főnév paradigmatikus rendszere* [The paradigmatic system of Hungarian noun]. Bp: Akadémia

[10] Sherwood, B.A. (1981): *Speech synthesis applied to language teaching.* In: Studies in Language Learning 3 pp 171-180

HUMAN FACTORS AND ACCEPTABILITY OF SYNTHETIC SPEECH AS AN INFORMATION SOURCE IN OPERATOR'S WORK

Vladimir Kuznetsov,
Laboratory of Experimental Phonetics
M.Torez Moscow Institute of Foreign Languages, Ostozhenka 38, Moscow 119034, USSR

Irina Frolova
Department of Applied Linguistics
Philological Faculty, Moscow University, USSR

Introduction

The real use of speech devices as partners of humans in communication act has revealed that traditional methodology of evaluating speech intelligibility does not provide a reliable prediction of the system performance in particular applications. In the present paper an attempt is made to investigate the role of human factors in perception of synthetic speech when listener is simultaneously engaged in other cognitive activities.

To simulate operator's work three types of task were used: (1) work with a switch-board; (2) solving of syllogistic reasoning tasks and (3) comprehension of short stories (jokes, parables) that involves deep semantic analysis and inference. In order to get a basis for comparison an identical experiment was carried out on speech material produced by human voice.

To enhance the diagnostic power of the experiments the synthetic voice (SV) and human voice (HV) were presented under two conditions: noise-free speech and speech masked by white noise (+N). The growth of subject's tiredness was checked by testing regularly STM (short-term-memory) capacity and WI (word intelligibility).

Methods and materials

According to the mode of speech generation and the presence of masking noise in the speech signal four experiments were designed and carried out in the following order: HV, SV, HV+N, SV+N. Each experiment was divided into three parts differing only in the presentation order of the experimental tasks. The total duration of the experiment varied from 1.5 to 2.0 hours.

STM capacity was measured before and after each experiment. The measuring procedure was as follows : the subject was presented twice with seven series of different figures; the number of figures in the series increased successively from 4 to 10; after presentation of each series the subject put down the figures he was able to recall [1].

WI was tested before the first and second parts of each experiment and after the third part. In WI tests the tables designed at Leningrad University were used.

Experimental material was tape-recorded and presented to the subjects through a pair of loud-speakers in an ordinary room.

Signal-to-masking noise ratio was +23 db. The ratio was determined experimentally: if the noise level was further increased, the intelligibility of SV fell sharply. The synthesis of speech material was performed on a "Phonemophon"-synthesizer, devoloped by B.M.Lobanov in Minsk.

Three subjects (one female and two males) took part in the experiments. Before the experiments the subjects were trained to acquire the de-

sired level of efficiency in performing the experimental tasks and to get familiar with the synthetic speech. During the training period the background data on the STM capacity of the subjects were obtained. After each experiment the subjects were interviewed.

Experimental tasks simulating operator's work

Work with a switch-board. By a command (for example: "Push a red plug into the socket A-5") the subject has to choose among the plugs of five different colours the appropriate one and put it into the socket as instructed. Positions of the sockets on the switch-board were specified by 12 letters of the Russian alphabet on the horizontal coordinate and by 16 figures (from 1 to 15) on the vertical coordinate. The interval between successive commands was 5 sec. Ten commands were presented at a stretch. To quantify the performance of the subjects the colour and position of the plugs were checked.

Syllogistic realoning tasks. The tasks were of the following type: "A is greater than B, B is greater than C, A ? C". Ten syllogisms were presented to the subjects in each part of the experiment. The interval between successive tasks was 5 sec. The presented syllogisms and the subjects' responces were tape-recorded. Thus, it was possible not only to detect subjects' errors, but to measure their reaction time as well.

Comprehension of short stories. In each part of the experiment the subject heard two or three texts. Their duration varied from 20 to 30 sec. After each text a male speaker read a list of questions constructed to examine the degree of text understanding. All questions, except the last one, were dealing with the surface semantic representation of the text. To answer the last question the subject was supposed to interpret correctly the deep semantic relations.

Results and discussion

Work with a switch-board. The subjects made no mistakes when the information was delivered by HV. In the experiments with SV the summary distribution of errors was as follows:

	1-st part,	2-nd part, of the experiment	3-d part
SV	1	4	9
SV+N	4	7	15.

Most of the errors were due to the low intelligibility of the letter designation of the horizontal coordinate. As far as the wrong selection of the plug's colour is concerned the subjects believed that they usually perceived the words designating the colours correctly, but could not recall them when it was necessary. This finding supports D.Pisoni's view that synthetic words perceived earlier are driven away from STM by subsequent speech events [2]. It should be noted that in the experiment "SV+N" the subjects showed signs of time deficit.

Syllogistic reasoning tasks. Table 1 presents the individual data on the error rates observed in the experiments. The slash separates the total number of errors from the number of cases when the subject could not solve a syllogistic problem. In these cases the subject's reaction time was considered to be 5 sec.

Table 1

	HV	HV+N	SV	SV+N
subj 1	5/2	1/1	7/5	7/4
subj 2	4/1	2	2	7/5
subj 3	0	0	4	3/1

There was no systematic correlation between the number of errors and the length of the subject's work in the experiment.

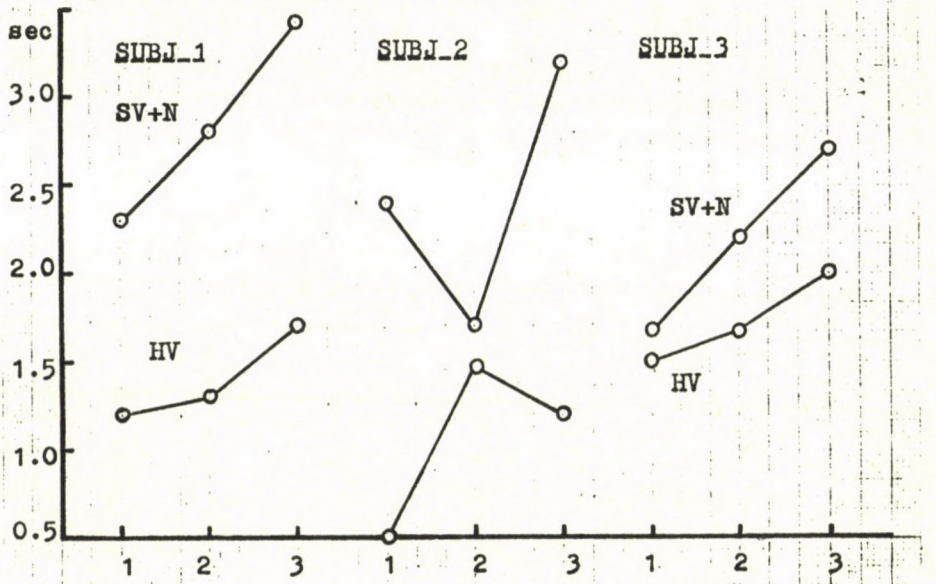


Fig. 1. Median reaction time for syllogistic problems.

Fig.1 displays median values of the subjects' reaction time as a function of the duration of work in the experiment (first, second and third parts). The data presented in Fig.1 concerns only two experiments: "HV" and "SV+N". There was the largest and most systematic difference in the values of reaction time obtained in these experiments. As a rule, reaction time varied considerably: the range of scatter reached 2.0-2.5 sec. Since the number of measurements was rather small and the individual differences were quite large, only tentative conclusions are possible. The most obvious conclusion to be drawn from Fig.1 is that the presence of masking noise and the length of work are significant variables affecting the operator's performance when he deals with synthetic speech.

Comprehension of short stories. Under all experimental conditions semantic interpretation of texts was adequate.

Measurements of WI and STM capacity. As it was expected, WI was perfect in the experiment "HV". Fig.2 depicts the data on the rate of confusions in the other experiments. Examination of Fig.2 shows that such factors as mode of speech generation and masking noise produce negative effects on the ability of subjects to recognize words.

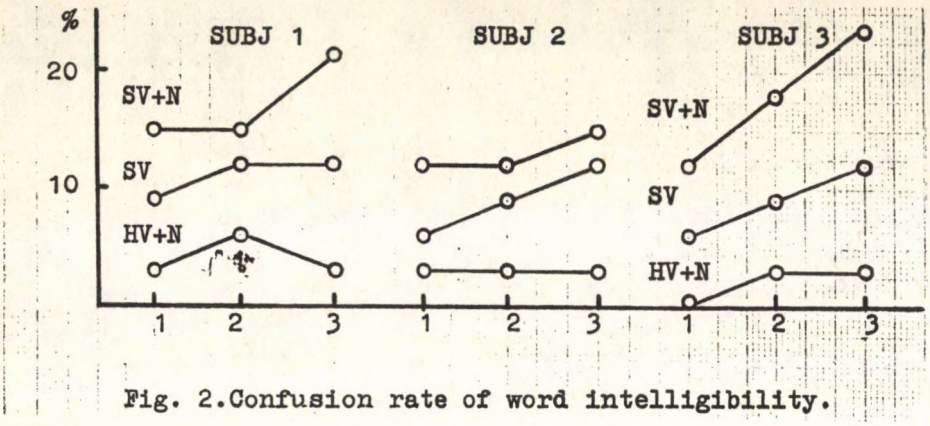


Fig. 2.Confusion rate of word intelligibility.

Table 2 lists the results of measuring the subjects' STM capacity before (B) and after (A) the experiments.

Table 2

	subj 1 B/A	subj 2 B/A	subj 3 B/A
HV	7.5/6.5	7.0/6.5	9.0/9.0
HV+N	7.0/6.0	7.0/6.5	8.0/8.0
SV	6.5/6.0	6.0/5.5	8.5/7.5
SV+N	6.0/5.0	6.0/5.5	7.5/7.0

As it is apparent from Table 2 the STM capacity consistently diminishes from the experiment "HV" to the experiment "SV+N". A lesser effect is produced by the workload in the experiments.

Conclusions

The use of synthetic speech has produced a negative effect on the efficiency of the subjects' performance in all experimental tasks, except comprehension of short stories. Even a slight masking of the investigated synthetic speech brings about appreciable increase in reaction time and errors. Fatigue and loss of efficiency progress faster when information is delivered by a synthetic voice.

References

1. Woodworth, R.: Experimental psychology.New York, Holt, 1938.
2. Pisoni, D.B.-- Nusbaum,H.C.-- Greens,B.G.:Perception of synthetic speech generated by rule. Proc. IEEE73. 1985, 1665-1676.

EMBEDDING SPEECH SYNTHESIS INTO APPLICATIONS

Géza NÉMETH, Géza GORDOS, Gábor OLASZY*, Attila TIHANYI
Speech Research Laboratory, Technical University of Budapest, Hungary
*Linguistics Institute, Hungarian Academy of Sciences, Budapest,Hungary

Summary

Lately one can see a dramatic increase in the number of products employing speech synthesis techniques. As many applications require limited vocabulary speech synthesis (LVSS), a fast, efficient, easy-to-use coding method is required. Such a system can be easily implemented by using a text-to-speech (TTS) system as the basis of code generation. An implementation of this approach, the self-contained FLEXIVOX LVSS development system, based on the multi-lingual SCRIPTOVOX TTS system, is presented.
Keywords: speech output, limited vocabulary speech synthesis.

Introduction

Until recently the results of speech synthesis research were rarely used in applications. So it was acceptable that speech researchers would be deeply involved in the design and the follow up of the more or less demonstration-like products. As new application areas emerge, their specific requirements must be taken into account. It is not very easy for researchers to fit their systems to these requirements. That is why there is a need for systems which can be easily operated by non-specialists with minimal support from experts. In this paper the general requirements for the application of speech output devices will be examined , along with the possible solutions for them; and a description of the FLEXIVOX LVSS development system, which can be regarded as a good solution for several classes of applications, will be given .

Considerations in applying speech output

The reason for the application of speech synthesis devices is that after a certain trigger event, or events, we want to receive a clear, understandable voice response. This response cannot be the same as that of a person, of course, but from a certain aspect the machine solution can approximate the human one. The major categories of classification are the following:

1) Number of messages
   In case of LVSS sytems we have a limited number of predefined messages while TTS systems provide an unlimited vocabulary.
2) Language dependence
   Some applications (e.g. airport announcement) require messages in several languages. In other cases (e. g. telephone access banking information sytems) one language is sufficient.
3) Speech quality

This is the parameter which is the most difficult to quantify. In most cases people replace it with speech intelligibility. This is a crude replacement as several other factors also determine quality.

4) Complexity of coding and replay
In case of LVSS, coding of speech parameters is necessary while in TTS systems this procedure can be omitted. High complexity can be the source of of quality degradation.

5) Interface, hardware implementation
It is very important for users (often system integrators) to be able to communicate easily with speech output devices. The most widespread interfaces are:
-industrial (contact closure, relay)
-numeric or ASCII keyboard
-instrument and computer interfaces (IEEE 488, RS 232, etc.)
The hardware implementation should be simple enough that production and service could be easily solved.

6) Price
As the speech interface is often an optional and/or new addition it is especially sensitive to the price/performance ratio.

For a given application one has to find an optimal compromise between the often contradictory requirements of the categories listed above. If intelligibility and low cost are the dominant factors, formant synthesis can provide a good solution in an integrated framework for both LVSS and TTS systems.

The FLEXIVOX system

Usually limited vocabulary and TTS synthesis methods and devices are developed on a different basis. To reach more human sounding speech output, LVSS systems take real speech as the source of information. Human speech is digitized (compressed) and stored for future retrieval. TTS systems on the other hand use language specific rules for converting a character string into speech. They are often accused of sounding robot-like, impersonal, etc.

However, recent development in TTS systems has led to devices with very high intelligibility scores . So if intelligibility is the major requirement in a LVSS application a TTS system can be used for the generation of the vocabulary elements. It is worth mentioning that a recent study showed that although subjects judged low bit rate LPC (2.4kbits) and ADPCM (9.6kbits) more pleasant than a TTS formant synthesizer, the latter reached better intelligibility results (5). The formant TTS based LVSS system offers great advantages in respect to both bit rate and ease of vocabulary generation and manipulation.

The FLEXIVOX system consists of the following components (the type of the microprocessors and/or the computers on which they are implemented is given in brackets):

Software

1) SCRIPTOVOX multi-lingual formant TTS system (Z8Ø, 65Ø2, I8Ø86, IBM PC, C-64)
The Hungarian version of the system is described in detail in (3,4). The other versions implemented (Russian, German, Esperanto, Finnish + Italian, Spanish under development) are based on the same structure. The input is an ASCII character string in the given language. The output is the formant synthesizer control codes provided by the rule

system. In this application the quality of the speech generated can be enhanced further with proper character string input. Typing in a few different versions and listening to them, one can choose the best solution. A regular intelligibility test carried out with the Hungarian version is described in (1).

2) Text editor (I8Ø86, IBM PC)
The length of a single message converted by the TTS system can be as long as 255 characters. The editor provides a limited set of commands for modifying them.

3) Formant synthesizer code editor (I8Ø86, 65Ø2, IBM PC, C-64)
This program offers sophiscated tools for editing the numeric values of the formant synthesizer's parameters. All values are displayed in meaningful units (formant frequencies, bandwidths, pitch in Hertz, frame length, amplitude in relative units). After some practice any intelligent user can modify the speech codes provided by the TTS system to reach better quality, the desired intonation contour.

4) Vocabulary editor (I8Ø86, 65Ø2, IBM PC, C-64)
This editor provides the possibility of inserting a message generated by the TTS system and the code editor into any vocabulary file. We can add a string of 16 characters to each message. Vocabularies can be initialized, filled or deleted with this program. One can exchange messages between vocabularies, concatenate messages of a given vocabulary for replay, etc..

5) Driver program for the MINIVOX production speech output card (Z8Ø, IBM PC)
If we burn this program into EPROM together with a vocabulary and place the EPROM into the MINIVOX card, it provides a complete LVSS output system with RS 232 interface. Speech output can be triggered by sending consecutively the code of the desired messages to the card. It contains a 1ØØØ character buffer. So it is possible to fill it during continuous speech output.

6) EPROM programmer program (I8Ø86, IBM PC)
This program can be used to burn a vocabulary and/or the MINIVOX driver program into EPROM. The 2716-27256 series are supported. It is also possible to use the program for reading back a vocabulary from EPROM. During programming it computes automatically the number of IC-s needed by a given vocabulary, checks their blankness and prompts the user to insert the IC to be programmed.

Hardware

1) FLEXIVOX development speech output card (IBM PC, C-64)
This card is coupled to the host computer's bus. It contains the formant synthesizer chip (and sometimes the language dependent databases).

2) SCRIPTOVOX TTS output card (Z8Ø, IBM PC)
It is a a self-contained Z8Ø based output card with RS 232 interface. The message to be converted can be as long as 1ØØØ ASCII characters closed by a point, exclamation mark or question mark. During conversion and speech output it is not possible to fill up its' buffer. It draws power either from the PC-bus or from a separate power supply.

3) MINIVOX production speech output card (Z8Ø, IBM PC)
This card can be used with the driver program described above or with proprietary software. It can also be powered either from the PC-bus or from a separate power supply. Correspondingly it can be used as a standalone unit or as a plug-in card for the PC. The card is equipped

with RS 232 interface.

4) EPROM programmer card (IBM PC)
This card is used in conjunction with the program described above.

5) Host computer (IBM PC with CGA monitor or C-64)
The C-64 based system provides somewhat limited possibilities at very low cost while the IBM PC version offers a complete self-contained solution for a wide range of TTS or LVSS applications.

A comparison of the traditional LVSS human input coding method and our approach can be found in (1).

Applications

As our system was implemented on three popular microprocessors (I8Ø86, Z8Ø, 65Ø2), it can be easily adapted to practically any application environment. On the C-64 it is used as a teaching aid (2) and as a speaking module for blind people. The Z8Ø TTS version was coupled to a character reader device and has been successfully tested for nearly two years. The applications of the LVSS system included output for Automatic Test Equipment (ATE), a talking blood-pressure meter, FDM communication measuring set, etc. In all of the LVSS applications coding was based on the TTS system. We had a very positive feedback from end-users concerning intelligibility. Most of the complaints related to naturalness and the limited number of voices.

Conclusion

After an overwiew of the factors which must be taken into account for the successful application of speech synthesis, one can see that the self-contained FLEXIVOX LVSS development system offers great advantages over the traditional human input method for fast, low-cost vocabulary generation when price and intelligibility are the dominant factors for the application.

References

1. CZAPP, L., GORDOS, G., NÉMETH, G., OLASZY, G., TIHANYI, A. : An integrated approach to text-to-speech and fixed vocabulary formant synthesis. Proceedings of the VDE-ITG Conference on Digital Speech Processing. Bad Nauheim 1988, 213--2216.
2. KOUTNY, I.--OLASZY, G. : Teaching Hungarian to foreigners by a talking computer. Proceedings of Speech Research '89. Budapest. In this volume.
3. OLASZY, G.--GORDOS, G. : On the speaking module of an automatic reading machine. Proceedings of the European Conference on Speech Technology. Vol. I. 1987. 225--28.
4. OLASZY, G. : A phonetically based data and rule system for the real-time text-to-speech synthesis of Hungarian. Proceedings of the Xth Int. Cong. of Phon. Sci. Utrecht 1983. 225---23o.
5. NIXON, C. W., ANDERSON, T. R., and MOORE, T. J. : The perception of synthetic speech in noise. Armstrong Aerospace Medical Research Laboratories Report, Wright-Patterson AFB, OH. 45433

SPEECH SYNTHESIS IN HUNGARY FROM THE BEGINNINGS UP TO 1989

Gábor OLASZY
Phonetics Laboratory, Institute of Linguistics
Budapest, Hungary

From time to time every scientific discipline has to survey the results of its history because the research of the future can only be effective if we rely on the past. The Speech Research '89 conference gives us a good opportunity to sum up the way of one branch of speech research in Hungary, namely that concerning the generation of fluent, unrestricted speech by machine for communication purposes*. This line of research is a special interdisciplinary one based mainly upon phonetics but involving several other disciplines as well, like theoretical linguistics, acoustics, mathematics, digital signal processing, computer technics, etc.

The development of speech synthesis in Hungary may be divided into three main periods: the beginnings, the research from the end of the 19th century and the breakthrough in the 80ies.

The beginnings

Almost every book or paper dealing with the history of speech research mentions the fundamental work in speech science of the Hungarian polyhistor Farkas Kempelen in the 18th century. He was the first scholar who systematically and scientifically observed and studied the process of speech production, the speech organs, the articulatory movements, the roles of the vocal cords, the mouth, the teeth, the tongue, and the lips when forming speech sounds. He did not rest satisfied with his theoretical findings, but continuously made experiments to construct a speaking device that can produce speech. He published the results of his 15-year-long work in the well-known book "Mechanismus der menschlichen Sprache" in 1791. Kempelen's speaking machine initiated the research of speech synthesis and deeply influenced the researchers of the following centuries (Wheatstone, Helmholtz, Riesz, Bell, Dudley, etc.)

Research from the end of the 19th century

Hungary had always been active in scientific research. By the second half of the 19th century a new branch -- called phonetics -- had begun to develop all over the world. A number of Hungarian linguists, physicists, engineers, and teachers joined this new direction already from the end of the 19th century (Gósy--Olaszy 1985) and were continouosly active in phonetics during the whole first half of the 20th century. Their work -- we know it in retrospect -- represents the foundation of modern theoretical and experimental phonetics including speech synthesis as well. I try to follow a chronological order summarizing their work.

The earliest experiments to investigate the acoustic structure of Hungarian vowels were made by the physicist Gy. Kont in 1894, then N. Klug made observations concerning the acoustic projection of articulation in

*(In this paper I do not deal with the research of various speech response technics that use digitally stored and later compressed human voice for limited vocabulary speaking systems. See Gordos, G. in this volume!)

vowels and he indicated the presence of energy maximums as the building frequency elements of vowels (today we call them formants). Some years later, in 1906, a singing teacher T. Szőnyi published with musical notes, the results of his experiments concerning the acoustic building elements of Hungarian vowels. He characterised o,a,o,u with one frequency value (one note), the others with two. (In 1984 a control synthesis was carried out using Szőnyi's data and the results were very close to the quality of Hungarian vowels (Gósy--Olaszy 1985). Experimental measurements of voiced/unvoiced detection, and measuring of sound energy was made by the linguist Z. Gombocz in 1900. Physiological measurements were carried out some years later (1908) by the linguist J. Balassa and by Gombocz as well.

The problem of timing in speech, the question of physical duration ratio of the sounds in a linguistic system are very important in the speech process. Gombocz was the first (1909) who made duration measurements by means of a kymograph. He discovered that the duration of a vowel may be influenced by the length of the word in which the vowel occurs. More than a decade passed before duration research was continued by J. Balassa (1921).

Melody is also a very important element in speech production. The first measurements in Hungary on melody patterns of speech were carried out by the phonetician L. Hegedűs (1930). He used a kymograph and a lupe to count the frequency values of vocal cord vibration during speaking. His measurements were so correct that after 50 years some of his results still correlate with those rules that were used in the Hungarian speaking SCRIPTOVOX system (Olaszy 1987).

From the 40ies the physicist T. Tarnóczy began intensive investigations on vowels and some other speech sounds (1941,1948,1964); in the 60ies the Hungarian phonetician K. Magdics (1965) made systematical investigations concerning the acoustic properties of all Hungarian speech sounds.

The breakthrough in the 80ies

This decade is the decade of intensive development in phonetics in Hungary. The foundation of this breakthrough was laid at the newly reconstructed phonetics laboratory of the Institute of Linguistics from the middle of the 70ies. Intensive research has begun on the field of speech acoustics, articulation, and perception supported by a measuring laboratory equipped with modern instruments, among others an OVE III formant synthesizer and from 1979 with a PDP 11/34 computer. (The results have been continouosly published in the series of Hungarian Papers in Phonetics from 1978.) The first synthesizing experiments were carried out from 1980 using the interactive speech synthesizing program INBERE developed at our laboratory for analysis-by-synthesis research.(Kiss-Olaszy 1982). Using this new tool the fine acoustic properties of speech sounds and CV, VCV, VC sound combinations were determined for Hungarian (Olaszy 1981, 1982). A special philosophy for formant based speech generation has also been formulated and from 1981 several speaking systems have been developed (table 1.) The main point of our philosophy was that we designed a minimalized and optimalized amount of acoustic building units (ABU) as the basis of speech signal generation.

Table 1 Formant based speaking systems in Hungary

SYSTEM	YEAR	MAIN FEATURES	LANGUAGE	COMPUTER	NOTE
INBERE	1980	A developing system to design speaking systems	no limitation	PDP11/34 + OVEIII	For laboratory use
SZAMOK	1981	A number and mathematical operation reader up to 1 billion	Hungarian	PDP11/34 + OVEIII	The first synthetic speaker
VOXON	1982	A speaking system with phonetic input	Hungarian	"	phonetic research
RUSSON	1982	A text-to-speech system	Russian	"	"
HUNGARO-VOX	1983	A full automatic text-to-speech converter	Hungarian	SYSTER + VOX-08	for industrial use
FLEXIVOX DEV.	1983 -- 1988	A developing system to design speaking systems	no limitations	C-64/128, IBM PC + MEA 8000	In cooper. of Phon.L. and Techn. Univ.
FLEXIVOX LV.	1986 -- 1989	A word developing system based on text-to-speech for limited vocabulary speaking cards	Hungarian German Russian Esperanto Finnish Italian	IBM PC + MEA 8000 and from 1989 for PCP 8200	for industrial use
BASICVOX	1984	A speaking screen system helping blind people in programming BASIC	Hungarian	C-64/128 + MEA 8000	price:150$
MIKROVOX 64	1985	A full-text-to speech system, programmable in BASIC with writing the text to be pronounced	Hungarian Esperanto German	C-64/128 + MEA 8000	for common use and for the blind
READING MACHINE	1986	A printed text reading machine with the RECOGNITA optical scanner (SZKI) and a text-to-speech converter	Hungarian	Speaking box with IBM PC	experimental
SCRIPTO-VOX	1984 -- 1988	An intelligent full text-to-speech converter card for general use in IBM PC/XT,AT	Hungarian	IBM PC + MEA 8000/ PCF 8200	for common use: language teaching, dyslexia correction helping blind people etc.
MULTIVOX	1986 -- 1989	A multilanguage text-to--speech card for IBM PC/XT,AT	Hungarian German Esperanto Finnish Italian Spanish	IBM PC + PCF 8200	

The main feature of these ABU's was that each of them was designed with a much shorter duration than a speech sound. The new point of this design was that one ABU was not strictly the representative of a given speech sound or sound combination, but it was used while building the speech signal at all places of the speech wave where the desired acoustic content was equal or closely equal to the acoustic content of the ABU. To organize the working of the ABUs a matrix type rule system was developed which got the input data from the text--phoneme code converter. The first, experimental, Hungarian speaking systems were a number reader uttering any kind of numbers and mathematical operations as well (Kiss-Olaszy 1981), a phonetic input reader VOXON (Bolla 1982), and the UNIVOICE, a real text-to-speech converter for Hungarian (Kiss-Olaszy 1982). UNIVOICE seemed to have the best sound quality at that time so intelligibility tests -- the first of such tests on machine voice in Hungary -- were carried out for the objective evaluation of its voice quality (Gósy-Olaszy 1983). The scores of the test showed a relatively high ie. 84% correct understanding for words and 98% for sentences. The next development was the HUNGAROVOX system designed for industrial purposes (Kiss-Olaszy 1983). This system was the first full text-to-speech one for Hungarian (patented in 1983). It converted every kind of text, numbers, abbreviations, etc. in real time into speech and produced a high quality, human-like voice. In 1983 the MEA 8000 free programmable formant synthesizer appeared (Philips) so from this time the development was concentrated -- in cooperation with the Technical University of Budapest -- on adapting the former results to this low cost chip, to open the way for practical applications. Table 1 contains a summary of the process of this research and development from 1983 as well. The further development and refinement of our philosophy became a good basis for the beginning -- from 1986 -- of developing speaking systems not only for Hungarian, but other languages as well. Thanks to a multilanguage contrastive research, today a six-language speaking system seems to emerge from the Phonetics Laboratory of the Institute of Linguistics.

References

Kempelen, W.: Mechanismus der menschlichen Sprache. Wien 1791.

Hegedűs, L.: Magyar hanglejtésminták grafikus ábrázolása. Kísérletfonetikai tanulmány. Collegium Hungaricum Füzetek V. Wien 1930.

Tarnóczy, T.: A magyar magánhangzók akusztikai szerkezete. Kir. Magyar Pázmány Péter Tudományegyetem Általános Nyelvészeti és Fonetikai Intézete. Budapest 1941.

Magdics, K.: A magyar beszédhangok akusztikai szerkezete. Budapest 1965.

Magyar Fonetikai Füzetek (Hungarian Papers in Phonetics (ed.): Bolla, K. 1--18. Budapest 1978--1988.

Gósy, M.--Olaszy, G.: A gépi beszéd megértése. (Az UNIVOICE magyar nyelvű, azonos idejű számítógépes szövegszintetizáló rendszer percepciós vizsgálata) NyK. 85. 1983/1, 93--104.

Gósy, M.-Olaszy, G.: A magyar fonetika első évtizedei. (A contribution to the early history of Hungarian phonetics) NyK. 87. 1985/1, 109--121.

Olaszy, G.: A magyar beszéd leggyakoribb hangsorépítő elemeinek szerkezete és szintézise. NytudÉrt. 121. Budapest 1985.

Olaszy, G.: Fonetikai alapú szabályrendszer a magyar beszéd automatikus, gépi előállításához. (Dissertation) Budapest 1987.

Olaszy, G.: Elektronikus beszédelőállítás. Budapest 1989.

IMPLEMENTATION OF ESTONIAN TEMPORAL STRUCTURE IN THE SPEECH SYNTHESIS-BY-RULE SYSTEM

Imre SIIL, Arvo OTT
Institute of Cybernetics
of Estonian Academy of Sciences

Speech synthesis is realized in this system on the principle of phonematic unit concatenation. There are 9 vowels and 19 consonants as basic units in the synthesis of Estonian speech. Coarticulation is modelled as a supplement using production rules.

The basic units are given for the formant model of the vocal tract controlled in real time by 12 parameters (see /7/). In general, the 'phonematic unit' corresponds to a single unit of synthesis which consists of two parts: the onset (TF segment on Fig.1) and the (quasi-)stationary part (or in other words - the characteristic segment). During the latter one the control parameters remain relatively constant. The structure of the stops and the trill is more complicated. They are described as consisting of several separate synthesis units each with normal structure.

Phonematic units are described in terms formant and energy values, which can be supplemented with values for basic duration. The basic duration is given as the duration of the whole unit (thus, including that of the onset).

Changes of duration are considered in the present model as additional components of phonematic unit duration. Starting from this simple assumption and applying the method of multiple linear regressions, it is possible to construct a quite flexible timing model (see the general determination of it in /7/). This model is relatively easy to realize in the production rules system. The duration will be added to the stationary part of the unit, since onset can be manipulated separately in the system of coarticualtion rules.

In this realization the phonological descriptions are assembled into the knowledge base structured for multigraduated production system. The knowledge can be represented in both static form (in declarations) and procedural form (in production rules). Rules determining duration constitute a relatively independent part of the whole knowledge base in the synthesis-by-rule system. The changes in this part do not violate the work of other rule systems in the knowledge base. And in turn, after removing or adding rules to one or another part of the knowledge base, the operation of timing rules system is guaranteed. The rule translator and rule interpreter have been programmed in microprocessor ASSEMBLER and are the parts of the speech synthesis development system residing on the PC.

In most synthesizers the control parameters are calculated after about 10 ms. Considering the difference limens of human speech perception this seems to be sufficient accuracy for the speech timing model. Probably, the minimum value of just noticeable differences in duration of segments is for normal

native Estonian speech not smaller than 7-10 ms Eek has reported the Weber ratio as being not smaller than 5% even in the functionaliy very important region of durations /2/. However, if we take into account the effect of combining the different factors which define small durational changes, the duration element must be less than 10 ms. On the level of time rule representation in this model a duration element of 2.5 ms is used. The final duration element after the rounding procedure of the rule interpreter is 10 ms.

Duration is calculated for phonematic units in a string derived from preceding transformation levels. The string can contain stress markers (' after stressed vowel) and indicators of phonologically relevant quantity degrees. Traditionally, there are 3 quantity degrees in standard Estonian, for instance, SA'DA (orthographically: sada), KATA (kada), where A and T are in the first quantity degree (Q1); SA'_DA (saada), KA'T_A (katta) - A and T are in the second quantity degree (Q2); SA'__DA (saada), KA'T__A (katta) - in the third quantity degree (Q3). Hence, some predefined procedures can be applied to get supplementary information about the context and the position of units.

In the technological process of speech generation the basic duration will be assigned to the phonematic unit only when it is not characterized by rules in the knowledge base. It is necessary to strive for optimization of the rule forming and computing processes, and this is why the basic duration ought to coincide with one of the more frequent phonematic unit values. In this model the initial values are 80-90 ms for vowels and 40-160 ms for different consonants (Q2 respectively 170 and 160 ms and Q3 260 and 190 ms).

The selection of basic duration means specifying the rate of synthesized speech at the same time. Comparing the temporal parameters of this model with the natural speech used in some experiments; the conclusion can be reached that the speech rate in the model is close to the average one for informants in investigations by Eek and Liiv /1/, /4/. However, some structural proportions are different in this model. But on the whole, concerning collation of the data obtained from analysis of natural and synthesized speech, it is justified only on the level of general tendencies. The simulation in speech synthesis systems can include only some of the important factors forming speech temporal structure.

The inherent durations of vowels are represented by rules. The rule description language for temporal modelling is in principle analogous with that quite simple one used by us in describing text-to-phoneme rules (see /6/ and examples below). In the left hand side of a rule, from amongst the factors govering duration, the following are described: the quality of a consonant after a vowel with primary stress; the position of the vowel in the stressed syllable; the position of the vowel in the first speech takt (i.e. the primary stressed syllable and the next one or two syllables together, synonym of 'foot'). For instance, rule (1) assigns 90 ms to the stressed vowels A and õ before the nasals. Rule (2) assigns 70 ms to the same vowels if they are the first components of a

stressed diphthong.

(1) IF s3=A ! õ & s3<str-vowels & s2<nasals THEN 36*2.5
e.g. in words KA'MA, MÕ'NUS.

(2) IF s3=A ! õ & s3<str-vowels & s2<long-vowels !
overlong-vowels THEN 28*2.5
e.g. in words KA'I_MUD, NÕ'E_LUME; KA'I__MU, NÕ'E__LUDA.

s_i marks the unit in the string (i,k=4,3,2,1,0,-1,-2,-3,-4), 'str-vowels', 'nasals' etc. are the predetermined classes of units, < means 'is a member', & works as logical "and", ! as logical "or".

Considering inherent duration of vowels with primary (main) stress, the expected difference between two groups is observed. The duration of U I Ü E is approximately half of that of A Ö Ä O (the vowels above appear in increasing order of duration) and also õ (cf. /4/). The inherent duration of consonants is already reflected in basic values. The basic value for the voiced consonants is 80 ms, for the greater part of voiceless consonants it is 100-120 ms. The inherent duration is shown particularly clearly in the word-initial position and between vowels, and in the case of voiceless consonants, also in the second and third quantity degree.

It stands to reason that in some cases the obtained duration reflects the spectral deficiency and inevitable incompleteness of the synthesis unit. Most likely, this finds its expression in a longer than expected duration of unit. In addition, the relation between quantity degrees and spectral characteristics is not sufficiently investigated and adequately formalized. This is particularly prominent on the occasion of vowels. For instance, in the case of õ: in Q3 compared with Q1 the first formant becomes lower and the frequency of the third formant increases. These tendencies are not revealed consistently by Q2 vowels (cf. /5/ which contains also articulatory motivation).

In creating this model, priority has been given to the realization of speech takt with main stress, because within the framework of that one only all the quantity degrees can present itself. Other speech takts, in this model have been limited to the indispensable regulation of single segments but no futher modification is made inside of the whole speech takt. These simplifications, however, do not disturb the normal perception of synthesized speech, in practice. It is possible that the norms of perception which are set up automatically after a dominating takt are more flexible and have some kind of compensatory effects.

It may be assumed that the speech takt behaves as an isochronic unit in the temporal programming of Estonian. The data from natural speech show that to warrant the isochrony of speech takt in revealing the Q2 and the Q3, it is necessary to cut down the duration of the component segments both before and after the one carrying the quantity contrast /1/. However, this concernes more the segments following it. Besides, the dependence of the unstressed part of speech takt upon the quantity degree of the speech takt is particulary noticeable (the average data about two-syllable takts).

In this model the inversely proportional dependence

between the phonematic unit carrying quantity contrast and vowels in next unstressed syllables has been fixed. Durational ratios of second-syllable vowels in the CVCV words by different quantity degrees are $V_{Q1}:V_{Q2}:V_{Q3}=1:0.92:0.81$. Minimum shortening takes place in consonants after the Q2 and Q3 vowel.

To represent the quantity degrees of a segment without disturbing the correct temporal proportions of the different parts of speech, it is necessary to use several phonetic parameters simultaneously (see also /1/,/3/,/4/). In this realization every quantity degree is connected to a fixed trajectory of pitch movements (Fig. 2). The present Q3 duration can be decreased on the account of additional energy impulse for the region where the Q3 segment begins.

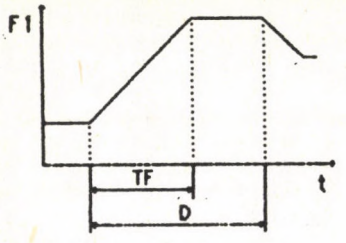


Fig.1

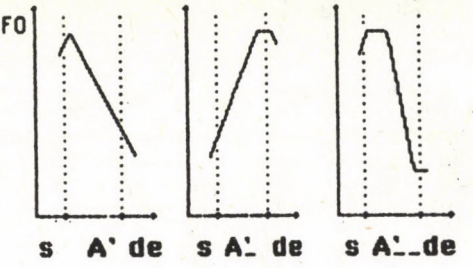


Fig.2

References

1. EEK, A.: Observations on the duration of some word structures: I-II. Estonian Papers in Phon. 1974,18--31; 1975,7--55.
2. EEK, A.: Just-noticeable differences of duration and language type: some preliminary notes. Estonian Papers in Phonetics., 1978, 21--26.
3. LEHISTE, I.: Experiments with synthetic speech concerning quantity in Estonian. The Ohio State University Working Papers in Linguistics. 9, 1971, 200--217.
4. ЛИИВ, Г.: Ударные монофтонги эстонского языка. Автореферат дисс. на соиск. ученой степени к.ф.н. Таллинн, 1962.
5. LIIV, G.: Acoustical features of Estonian vowels pronounced in isolation and in three phonological degrees of length. ENSV TA Toimetised. XI, Ühisk.-teaduste seeria, 1, 1962, 63--97.
6. OTT, A.--SIIL, I.: Real-time speech synthesis - development and employment. Computers and Artificial Intelligence. 6, 1987, 173--180.
7. OTT, A.--SIIL, I.: The synthesis-by-rule development system with expert capabilities. Proc. of the 11th Int. Congr. of Phonetic Sciences, 3. Tallinn, 1987, 278--281.

ENTWICKLUNG VON REFERENZMUSTERN FÜR AUTOMATISCHE LAUTBEURTEILUNGEN

Eberhard STOCK, Uwe HOLLMACH, Frauke SUCKOW
Wissenschaftsbereich Sprechwissenschaft Sektion
Germanistik/Kunstwissenschaften der Universität
Halle, Halle, DDR

Einführung

Die Verfügbarkeit von preiswerten und relativ leistungsstarken Kleinrechnern provoziert auch für die sprechwissenschaftlich-phonetische und logopädisch-phoniatrische Forschung, Lehre und Therapie die Frage, wie diese Rechner als Arbeits- und Rationalisierungsmittel eingesetzt werden können. Die internationale Literatur weist aus, da3 entsprechende Überlegungen in mehrerlei Richtung angestellt worden sind. Das betrifft z.B. die Erfassung der Leistungsparameter der Stimme mittels des sogenannten Stimmfeldes und des Sängerformanten oder die Darstellung der Artikulationsbewegungen für eine eingegebene Phonomsequenz oder die automatische Beurteilung von Phonem- und Intonemrealisierungen.

Für das phonetische Training im Fremdsprachenunterricht, für die Übungsbehandlung von Sprachstörungen und für die Tauglichkeitsprüfung bei Bewerbern für sprechintensive Berufe (Lehrer, Schauspieler, Rundfunk- und Fernsehsprecher) ist die durch den Lehrenden oder Therapeuten vorzunehmende Beurteilung von Phonemrealisationen nach dem Kriterium der Korrektheit ein überaus häufiger Routinevorgang, der nach Rationalisierung verlangt. Eine solche Rationalisierung mit Computer bietet au3erdem den Vorteil, da3 durch die Objektivierung auf dem Monitor der Lernende visuell und folglich leicht erkennen kann, wieweit seine Artikulation mit dem Muster übereinstimmt. Diese Erkenntnis kann sich zweifelsohne auf seine Motivation günstig auswirken.

Die Beurteilung von Realisationen der Phoneme s und z stellt dabei aus zwei Gründen ein Sonderproblem dar: 1. Mehrere hallesche Untersuchungen weisen aus, da3 die Häufigkeit von unkorrekten S-Realisationen (Sigmatismen) zunimmt und da3 vor allem Sprecher mit einem sogenannten apikalen S (die Zungenspitze schwebt frei hinter den oberen Schneidezähnen) zu einem Sigmatismus tendieren. Diese Entwicklung könnte Einflu3 haben auf die Kodifizierung des Aussprachestandards im Deutschen, der in unserem Aussprachewörterbuch beschrieben wird (1). 2. Die Beurteilung von S-Realisationen wird für Personen, die älter als 50 sind, mit zunehmendem Alter immer schwerer, weil die Hörfähigkeit für höhere Frequenzen nachlä3t und die Brillanz bzw. Schärfe von Geräuschen nicht mehr eingeschätzt werden kann (3). Die Feststellung von korrekten oder unkorrekten S-Artikulationen bedarf deshalb einer zusätzlichen Kontrolle.

Diese beiden Gründe sind für unsere sprachkulturell orientierte Arbeit sehr wichtig. Sie haben uns veranla3t, die Untersuchungen, zur Entwicklung von Referenzmustern für automatische Lautbeurteilungen mit der Aufstellung von entsprechenden Mustern für die S-Phoneme zu beginnen. Die folgende Darstellung ist daher nur als Beispiel zu verstehen, und sie beschränkt sich auch nur auf ein Problem, nämlich auf die Realisation des /e/ im Wort Kies. Die Probleme der Kontextabhängigkeit und der phonostilistischen Variation werden hier nicht beachtet.

Methoden

Von F. Suckow wurden 80 Probanden gewonnen, die einen 20-Zeilen-Text und 8 Testwörter vorlesen mu3ten. Die entsprechenden Tonbandaufnahmen wurden 7 Experten im Alter zwischen 20 und 30, die durch Hochfrequenzaudiometrie (2) als hörgesund und für S-Untersuchungen als tauglich befunden worden waren, vorgeführt, wobei folgende Fragen zu beantworten waren: 1. Beurteilen Sie die S-Laute als korrekt? 2. Sind die unkorrekten S-Laute zu scharf, zu stumpf, oder ist ihr Schärfegrad nicht zu bestimmen? 3. Ist die Normabweichung sehr gering, gering oder stark? 4. Welcher Sigmatismusform müssen sie die unkorrekte Lautbildung zuordnen? 5. Entspricht die Qualität der S-Realisationen den Anforderungen für Lehrer? 6. Entspricht die Qualität der S-Realisationen den Anforderungen für Sprecher in den elektronischen Medien?

Die abgegebenen Urteile wurden methodenkritisch geprüft und die in den Antworten erkennbaren Divergenzen und Widersprüche nach ihrer Beurteilungstendenz bewertet (5). Für die Aufstellung des Referenzmusters wurde die S-Realisierung im Wort Kies von 41 Sprechern (21 weiblich, 20 männlich) herangezogen. Es handelte sich um diejenigen Probanden, deren Lautbildung mehrheitlich als korrekt beurteilt oder mehrheitlich als tauglich für den Lehrerberuf bzw. für das Mikrophonsprechen empfunden wurden.

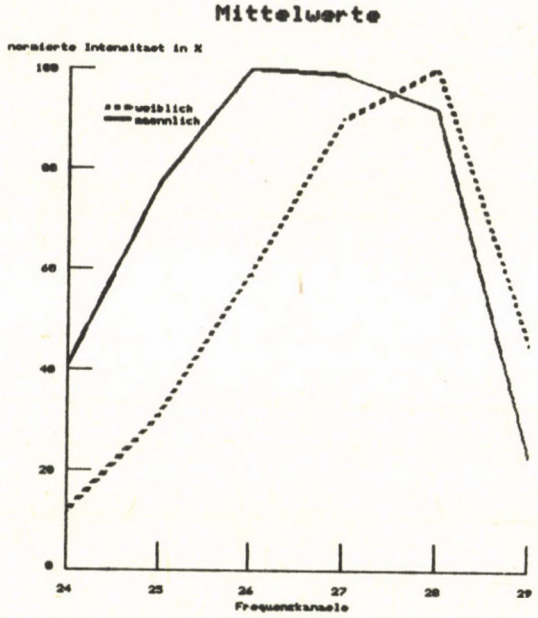
Für die Spektralisierung und Digitalisierung der fraglichen 41 S-Allophone wurde eine von Hollmach gebaute computergesteuerte Filterbank mit 32 Kanälen benutzt. Der Frequenzbereich liegt zwischen 80 Hz und 20,5 kHz; für flie3endes Sprechen ist ein Zeitfenster von 6 ms einstellbar; der Abstand der diskreten Filterausgangssignale beträgt eine Viertel Oktave. Es wurde au3erdem ein Analog-Digital-Umwandler mit 10 bit Auflösungsvermögen (entspricht 1024 Stufen) verwendet (4). Die Segmentierung wurde an den Farbdisplays der gewonnenen Computer-Sonagramme ohne Schwierigkeit durch das manuelle Setzen von Lichtbalken vorgenommen.

Die Entwicklung der Referenzmuster geschah auf folgendem Wege: 1. Durch das Setzen einer Bandpa3filterung wurden die zu bewertenden Frequenzen auf den Bereich von 0,99 bis 18,84 kHz (Mittenfrequenzen der Filterbankkanäle 15 und 32) eingeengt. 2. Um die individuellen Lautheitsunterschiede auszugleichen, wurden die digitalisierten Frequenz-Intensitäts-Strukturen normiert, indem für jedes S-Allophon der Filterausgangswert mit dem höchsten Betrag als 100 % gesetzt und alle anderen Filterausgangswerte ihrer Relation entsprechend automatisch prozentual bewertet wurden. 3. Die auf diese Weise entstandenen Kurven der normierten Filterausgangswerte wurden zunächst subjektiv bezüglich ihrer Kurvenform (Intensitätsverteilung) sowie ihrer Frequenzanteile (Beteiligung der Filter) überprüft. Es zeigte sich, da3 die Kurvenformen unterschiedlich ausfielen. 4. Deshalb wurden die einzelnen Kurven zum Geschlecht der Sprecher und zu den Expertenurteilen hinsichtlich des Schärfegrades der S-Geräusche in Beziehung gesetzt. 5. Für die S-Allophone der weiblichen und der männlichen Sprecher wurden die normierten Filterausgangswerte getrennt arithmetisch gemittelt. Als Streuungsma3e wurden die Varianz und die Standardabweichung für jeden Filterausgang berechnet.

Ergebnisse

Gegenüber der Mittelwertkurve für die männlichen Sprecher ist die der weiblichen Sprecher in die Höhe verschoben. Der mit 100 % bewertete Filter

ausgangswert für die Männer fällt auf den Filterkanal 26 (Mittenfrequenz 6,7 kHz), der für die Frauen auf den Kanal 28 (Mittenfrequenz 9,4 kHz). Die grö3te Stabilität (kleine Streuungsma3e) haben bei den Männern die Filterkanäle 26 und 27, bei den Frauen 27 und 28 (siehe Abbildung).



Der Schärfegrad des S-Geräusches korreliert mit der Steilheit der Kurvenflanken. Für die automatische Bewertung wurden die Differenzen zwischen dem mit 100 % bewerteten Filterausgang und den beiden benachbarten tiefernliegenden Filtern sowie mit dem 100% Wert und dem benachbarten höherliegenden Filter summiert. Lag diese Summe zwischen 120 und 170, so handelte es sich um ein als überscharf, aber korrekt beurteiltes S-Allophon. Die günstigsten Expertenurteile wurden für S-Allophone abgegeben, die eine entsprechenden Summe zwischen 30 und 80 aufwiesen, die also Kurven mit relativ flacher Steigung hatten

Diese Mittelwertkurven, die nach unserer Auffassung als Referenzmuster fungieren können, wurden dadurch auf ihre Tauglichkeit für die automatische Lautbeurteilung getestet, indem eine Reihe von sigmatischen S-Allophonen (Sigmatismus stridens, Sigmatismus addentalis, Sigmatismus interdentalis, Sigmatismus lateralis) mit den Kurven verglichen wurden.

Der Sigmatismus stridens ist dadurch gekennzeichnet, da3 die als Indikator für die Kurvensteilheit fungierende Summe über 170 liegt. Der Computer kann folglich durch eine einfache Berechnung entsprechende S-Allophone automatisch erkennen und abweisen.

Die Sigmatismen addentalis, interdentalis und lateralis werden durch den Rechner zweifelsfrei dadurch erkannt, da3 ie im Bereich von 990

bis 2350 Hz (Filterkanäle 15 bis 20) Intensitäten aufweisen. die bei den korrekt beurteilten S-Allophonen in jedem Falle fehlen. Diese zusätzlichen Intensitäten bei den genannten Sigmatismusformen sind offensichtlich verantwortlich für die Stumpfheit der entsprechenden S-Geräusche.

Diskussion

Die ermittelten Referenzmuster für die korrekten S-Laute bei weiblichen und männlichen Sprechern ermöglichen die automatische Abweisung von Sigmatismen und die Bewertung von S-Geräuschen auf ihren Schärfegrad hin.Sie gestatten es dem Lernenden, seine S-Allophone mit dem Muster zu vergleichen. Sie sind vor allem für diejenigen Sprachlehrer und Therapeuten eine Hilfe, wenn durch den altersbedingten Rückgang der Hörfähigkeit für hohe Frequenzen die Beurteilung der S-Allophone nicht mehr möglich ist. Wir halten es für denkbar, mit dieser Methode auch Referenzmuster für die automatische Beurteilung anderer Phonemrealisationen zu entwickeln. Mit solchen Referenzmustern kann das phonetische Training unterstützt und die individuelle Studienarbeit, die sonst unbedingt die Kontrolle des Lehrenden erfordert, erleichtert werden.

Literatur

1. KRECH, E.-M. et al.: Gro3es Wörterbuch der deutschen Ausprache. Leipzig, 1982.
2. HOLLMACH, U.: Der Vergleich metrischer und auditiver Daten von S-Allophonen im zusammenhängenden Sprechen. Diplomarbeit, Halle, 1984.
3. STOCK, E. und HOLLMACH, U.: Zur Identifizierung und Bewertung von S-Allophonen. Festschrift für H.-H. Wängler. Hrsg.: R. WEISS (Beitr. zur Phonetik und Linguistik, Bd. 52). Hamburg, 1987, 381--397.
4. STOCK, E. und HOLLMACH, U.: Objektive Bewertung von S-Allophonen. Proc. XIth ICPhS. Tallinn, 1987, Bd. 5. 423--426.
5. SUCKOW, F.: Zur Realisation von S-Lauten. Phil. Diss. A, Halle, in Vorbereitung.

SPEECH RECOGNITION SYSTEMS IN ACOUSTICAL RESEARCH LABORATORY

VICSI K, BERÉNYI P.

Acoustical Research Laboratory of Hungarian Academy of Sciences. Budapest, Hungary

INTRODUCTION

In the Acoustical Research Laboratory the digital speech processing and computer application for speech research were started at the end of 1984. The research work was done on the line of Tarnóczy and used up all advantages of digital technics. The aim was to develope speech recognizers. Our first speech recognizer got ready at the midle of 1987. That recognizer is a stand-alone, speakér-dependent eqipment. It handles any 80 words by 10 users. It was developed for the Electroacoustical Factory in Budapest.

The speaker dependent isolated word recognition system with middle size dictionary was ready at the end of 1987. This system is able to recognize some hundreds words on any of IBM PC. Next year the system was brought to perfection. The recognition time was decreased under 500 ms for 300 words. Reorganizing the program the voice substitutes the keyboard.

Nowadays continuous speech recognition system is under development, cooperating with Technical University. An other common topic is to recognize words under noisy environments.

AUDITORY MODEL

An auditory model was constructed for acoustical analysis of speech. In that model critical band filtering is used. The 20 filters have bandwiths of Bark /5/ . The operating frequency range is 80Hz to 8kHz. The filters have asymptotic slopes which reflect the filtering characteristics of the human ear. The slope of these filters is steaper towards high frequencies - 25 dB/Bark - than towards lower frequencies - 10 dB/Bark - corresponding to the masking curves /2/. After rectification the time delay /τ/ of the filter outputs decreases with the increasing center frequencies until 1kHz. This time delay is about five times longer then the periodic time of the center frequency. Above 1kHz the time delays are 5 ms. constantly. The Fig. 2. shows the change of loudness, energy below 1 kHz and zero crossing number in time. The lower part of the figure shows how the 20 filter outputs are changing in time, representing a sonogram-like from. On the figure the following phonetically balanced sentence fragment is presented: "A falatozo: ban sort, bort, ydito: ita". It was spoken by female.

Still 5,6 years before some researchers /1/ came to the conclusion that Bark filter bank is not a good analysing form for speech recognition, but nowadays more and more big research centers give preference to auditory model /4/ in their speech recognition systems.

ISOLATED WORD RECOGNIZERS

The block diagram of our small and middle size dictionary recognition systems are to be seen in Fig. 1. It is evident that an optimal parameter extraction makes the recognition accuracy as high as possible and decreases the recognition time as short as possible. In our systems every 10 ms frame are characterised by 5 parameters.

Further improvement of the systems could be by a good choice of pattern

SYSTEM CONFIGURATION

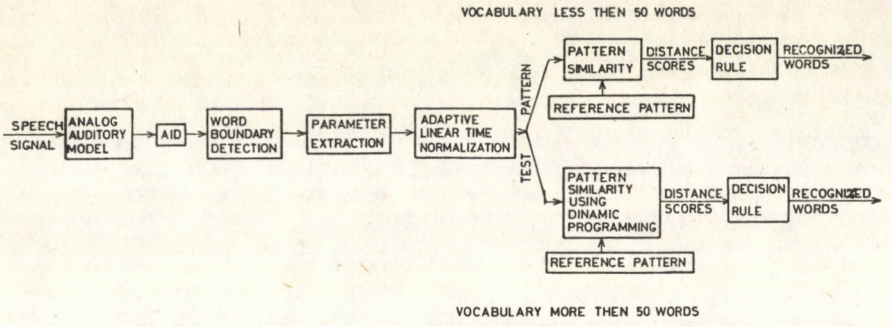


Fig. 1.
Block diagram of the isolated word recognition system

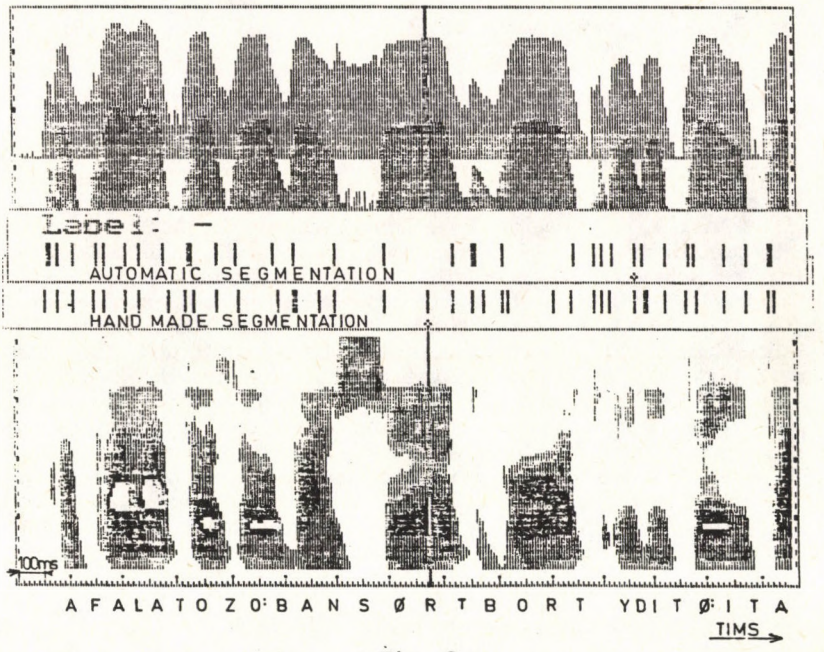


Fig. 2.
Automatic and hand-made segmentation of continuous speech

similarity measurement even if the world-wide known symmetric DTW algorithm is used /3/. For example, to find the optimal length of the adjustement window.

/R=2r/ is very important. If in the dictionary the words have the same syllable-number, the optimal R is about 80 ms. In that case when the dictionary is composed of one-, two-, tree - syllablic words the optimal R is 140 ms, at normal speaking rate. Three different ways of pattern similarity measurement were compared /100 words were in the dictionary/:

1. Words collected into groups according to the syllable-number /Grouping/ adaptiv-linear time normalization /ALTN/ inside the groups;
2. Grouping and ALTN inside the groups and DTW with R=80 ms;
3. ALTN without grouping and DTW with R=140 ms were carried out.

Table 1.

Pattern similarity measurement	Recognition Accuracy
1. Grouping, ALTN inside the groups	92%
2. Grouping, ALTN inside the groups and DTW with R=80 ms	97%
3. ALTN without grouping, DTW with R=140 ms	91,5%

As it is to be seen from the table the grouping is very useful before the traditionally used DTW. The recognition accuracy is much better, and the recognition time is shorter.

CONTINUOUS SPEECH RECOGNITION

Our recognition system, at first step, is developed to recognize the continuous speech only on phonetic level. Simultaneously data base for recognition of Hungarian language is constructed.

Following levels were used for the recognition: - acoustic analysis by the auditory model - segmentation - classification into larger phonemic categories - labelling /candidates are placed into a phoneme lattice/. Taking phonemic and phonological rules into consideration the decision of a lower level is corrected at a higher one.

The success of recognition results depends considerably on the success of segmentation. For the result not only the segmentation technique, but also the choice of the suitable phonemic units are important. The language, the aim of recognition, the preprocessing system and many other things determine which phonemic units are the best. The units of our segmentation are generally phonemes, but adapting to the acoustic features - obtained by auditory model - sometimes the segment is shorter, sometimes longer than the phoneme.

The rule based method for the segmentation was used and developed in many steps. The segment - baundaries were corrected step by step.

At the first step, a simple spectral change measurement is used:

$$\Delta \bar{E}_{i+1} = \frac{1}{20} \sum_{j=1}^{20} \left(E_{i+1}^{j} - E_{i}^{j} \right) / 2 \; ;$$

where E_i^j represents the log energy of the j-th filter outputs at i-th frame. The spectral change was more characteristic for boundaries when the i-th

frame was compare with the /i+2/-th frame, then with /i+1/-th one.

In the second step, the segments having been classified into phonemic units, the boundaries can be corrected using phonemic rules. For example, silent, or closure period of stops were decided when the total energy was under a mooving threshold. The problem is whether these segments are really silent or they belong to stops or affricates. It can be decided, if the following segment is examined. At those frames where the full energy is greater than the energy below 1 kHz burst and/or spirant noise must be present, and these are parts of stops or affricates. Taking several similar rules into consideration mentioned above, the automatic segmentation was done, and the results were compared with the hand-made ones. An example is shown in Fig. 2. This segmentation method works very well in nearly all cases of sound connections, except leterals, which are frequently connected together with vowels.

On the base of 10 phonetically balanced sentences /464 phonemes/ 42 boundaries were missed, all were laterals in VC or CV connections, and only 2 plus boundaries were found. Burst part was not calculated into that statistic, but all unvoiced stop-bursts were found in correct way, while voiced stop-bursts were found only in some cases.

For the labelling three different methods are used: the rule based method, the dinamic time warping and the method of Marcov modeling. By the comparison of the results of these methods we could decide which method is the best for our system, and how the whole data base must be constructed.

REFERENCES

/1/ Blomberg M., at al.: Auditory Models in Isolated word recognition. IEEE ICASSP 1984. 17. 9.1.-17.9.4.

/2/ Fourcin, A.J. at al.: Speech Processing by Man and Machine, "Group Report" on Recognition of Complex Acoustic signals, edited by T.H. Bullock, Life Sciences Research Report of the Dahlem Vorkshop, Berlin 1977.

/3/ Sakoe, H. - Chiba, S.: Dinamic Programming Algorithm Optimization for Spoken Word Recognition. IEEE Trans. on Acoustics, Speech and Signal Proc. /1978/ Vol. ASSP-26. No. 1. pp. 43-49.

/4/ Proc. of Montreal Symposium on Speech Recognition. Montreal 1986.

/5/ Zwicker, E.: Psychoakoustik, Springer Verlag, Berlin 1982.

DIE GRUNDPARAMETER DES MIKRORECHNERS FÜR DIE MESSUNG VON SUPRASEGMENTEN

Julius ZIMMERMANN
Das Phonetiklaboratorium, Die Philosophische Fakultät
der Pavol-Jozef-Šafárik-Universität, Košice, ČSSR

Die Sprachsignalverarbeitung mit Hilfe von diskreten Methoden wird praktisch von einem Zentralrechner, Mikrorechner, Signalprozessor oder einer spezialisierten Digitalanzeige realisiert. Die Anwendung eines Gro3rechners ist, was die Schnelligkeit und Genauigkeit betrifft, am günstigsten, zugleich aber auch am aufwendigsten. Spezialisierte logische Systeme, mit einem A/D-Umsetzer ausgestattet, sind wenig variabel, und ihre Entwicklung ist teuer. Die Anwendung eines Signalprozessors ist effektiv und perspektivisch, er ermöglicht aber nur in beschränktem Ma3e die Änderung der Systemeigenschaften und -konfiguration, weil seine Parameter schon bei der Herstellung eines konkreten Signalprozessortyps festgelegt werden. Die praktisch günstigste Anlage ist ein Mikrorechner. Im Hinblick auf die gegenwärtig massenhafte Verbreitung der Personalcomputer, die mit IBM PC/XT/AT kompatibel sind, ist ihre Anwendung begründbar, jedoch unter der Bedingung, da3 der Benutzer die unabdingbaren ergänzenden Karten gewinnt, wie einen A/D-Umsetzer, einen D/A-Umsetzer, einen Modul diskreter Ein- und Ausgaben, einen Modul der Serien- und parallelen Anschlu3stelle und gegebenenfalls auch einen Multiplexer zum ADU. Man nimmt an, da3 der Mikrorechner mit einem mathematischen Ko-Prozessor versehen ist.

Eine andere Möglichkeit stellt die Anwendung eines variablen Mikroprozessor-Baukastens dar, das zum Zusammenbau der Steuerungssysteme in den Anlagen anderer Finalprodukthersteller gebraucht wird (OEM). Das Blockdiagramm eines solchen Baukastens ist auf Abb. 1 zu sehen. Es ermöglicht einen Systembetrieb mit mehreren Prozessoren, den sog. Multimasterbetrieb.

Bei der Auswahl geeigneter Modultypen und deren Sortiments geht man von der Frage aus, welche Ansprüche an die Eigenschaften des Programms, das das Sprachsignal verarbeiten soll, und an die Bedienung anderer externer Zusatzgeräte (XY-Schreiber, Oszilloskop, Joystick, Verstärker usw.) gestellt werden. Es ist günstig, wenn der betreffende Arbeitsplatz über ein Entwicklungssystem mit einem Emulationsadapter verfügt, wodurch die Entwicklung und Fehlerkorrektur von Programmen beschleunigt wird.

Beim Bilanzieren der Geschwindigkeit mu3 man die Geschwindigkeit der A/D-Umwandlungen und des D/A-Umsetzers sowie die des eigentlichen Mikroprozessors auswerten, und das auch in dem Falle, wenn es sich um einen Mikrorechner mit DMA-Steuerung handelt. Die Sukzession der Eingabebefehle des Sprachsignals auf Assemblerbasis ist nach einer notwendigen Initialisierung des Umsetzers wie folgt:

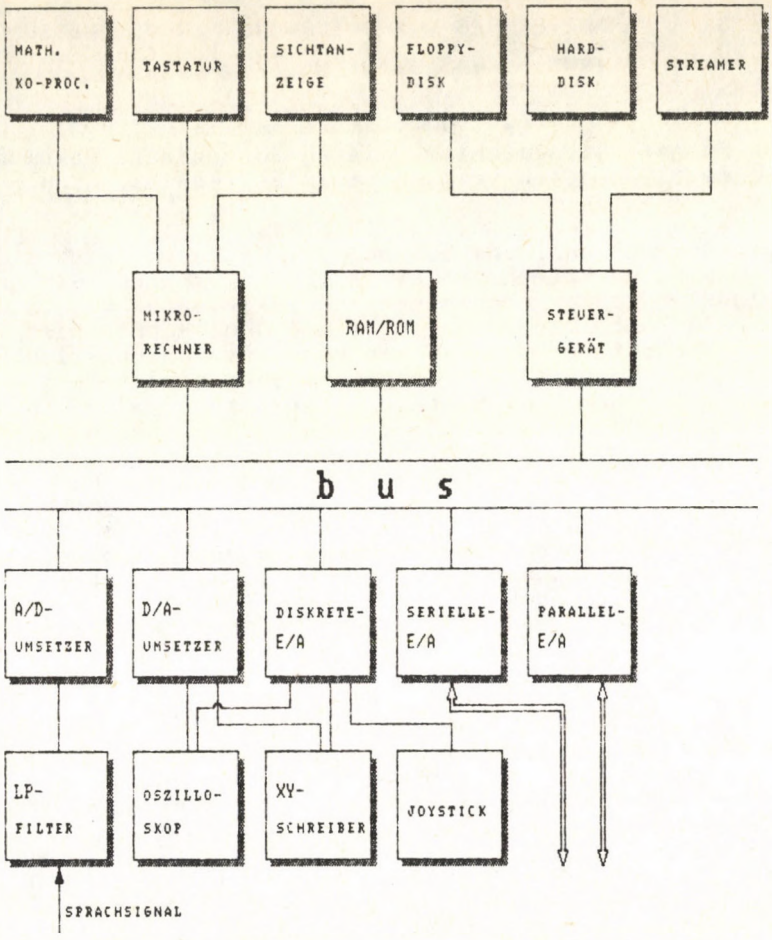


Abb. 1

```
1    Start:   LXI   B,ADRADC   ;Start der A/D-Umwandlung
2             LDAX  B          ;Übertragung der Stichprobe
:             :                ;in den Akkumulator
:             :
n             JMP   Start      ;Wiederholung der Umwandlung
```

Die Zeit der Umwandlung des A/D-Umsetzers beträgt Einheiten der Mikrosekunden. Der Sprung zum Befehl 1 - Start der Umwandlung kann erst nach dem Ablaufen der Umwandlungszeit vom vorangehenden Start programmiert werden. Ferner ist das

Shannon-Theorem zu beachten:

$$f_{smpl\ max} \geq 2 f_{sig\ max}$$

Diese Bedingung kann durch das Einsetzen vom LP-Filter auf die A/D-Umsetzereingabe erfüllt werden. Das Einsetzen eines entsprechenden Filters auf die A/D-Umsetzerausgabe ist notwendig, es sichert die Korrektur des stufenförmigen Ausgabesignals.

Mit Hilfe von Befehlen 2 + (n-1) kann die eigentliche Verarbeitung der Stichprobe eines Sprachsignals realisiert, externe Zusatzgeräte (Oszilloskop, Schreiber usw.) bedient und die Stichprobenzahl gezählt werden; es können die Stichproben im Operativ- und Externspeicher gespeichert werden, wenn es nicht um die Verarbeitung in der Echtzeit geht.

Bei der Verarbeitung des Sprachsignals in der Echtzeit, d.h. beim Produzieren des verarbeiteten Signals in der D/A-Umsetzerausgabe während der Abtastung mu3 zwischen Befehl 1 und Befehl n noch die Befehlsfolge für die D/A-Umwandlung eingefügt werden:

```
Ausg:   MOV     A,M     ;Speicherinhalt in den Akkumulator
        STA     ADRDAC  ;Akkumulatorinhalt in den DAU
         :
         :
        JMP     Ausg    ;Wiederholung der Umwandlung
```

Die Breite des aus dem A/D-Umsetzer ausgehenden resp. in den D/A-Umsetzer eingehenden Wortes gibt das Auflösungsvermögen für analoge Seiten beider Umsetzer an. Der 8-Bit-Umsetzer löst 256 Niveaustufen, der 10-Bit-Umsetzer 1024 Niveaustufen auf.

Der Mikroprozessortyp (8-Bit- resp. 16-Bit-Prozessor) stellt das Grundparameter bei der Zeitbilanz der Tätigkeit des ganzen Systems dar. Die Durchschnittszeit eines Befehls beim 8-Bit-Prozessor beträgt 2 - 8 Mikrosekunden, der 16-Bit-Prozessor kann 10-mal schneller sein.

Das aufgrund der angeführten Überlegungen vorgeschlagene Mikroprozessorsystem kann - ähnlich wie der Speicheroszillograph - zur Verfolgung des Redeoszillogramms, zur Aufzeichnung des Oszillogramms auf den XY-Schreiber, zur Segmentierung des Redeflusses, zur Verfolgung der Zeitdimensionen und zur darauffolgenden Berechnung der durch die Zeitmodulation der Stimme herausgebildeten Suprasegmente, zur Erforschung der Intensitätslinie und zur Entdeckung F_0 dienen. Aufgrund dieser Parameter kann mittels Software die Berechnung weiterer Parameter, z.B. die des Akzents erfolgen.

References

1. HESS, W.J.: Pitch Determination of Speech Signals. Springer - Verlag, Berlin, 1983.
2. NEY, H.: Bestimmung der Zeitverläufe von Intensität und Grundperiode der Sprache für die automatische Spreche-

erkennung. Frequenz 35 (1981), No. 10, s. 265-270.
3. RABINER, L.R.--GOLD, B.: Theory and Application of Digital Signal Processing. Englewood Cliffs New Jersey, Prentice Hall Inc. 1975.

A VERSMONDÓI STÍLUSRÓL TÖRTÉNETI MEGKÖZELÍTÉSBEN

ANTONI Andrea
Janus Pannonius Tudományegyetem
Pécs, Magyarország

Előadásomban a versmondói stílus változásának fonetikai vetületéről és szociokulturális vonatkozásairól számolok be. Az interpretatív stílus alakulását három, különböző generációhoz tartozó színművész versmondásában követem nyomon: Ódry Árpádéban (1876-1937), Latinovits Zoltánéban (1931-1976) és Mensáros Lászlóéban (1926-). Mindhárman Vörösmarty Mihály Vén cigány című versét adják elő.

A versmondói stílus változása vizsgálatának érdekében fel kell elevenítenünk néhány fontos ismeretet, hacsak vázlatosan is. Az elhangzó beszédet különböző stíluskategóriákra lehet bontani "hangzása, indíttatása, valamint a beszédhelyzetek alapján". Eszerint beszélhetünk élőszóról, felolvasásról, valamint a kettő közötti átmenetekről: reproduktív - interpretatív illetve fél-reproduktív beszédről. A számunkra fontos reproduktív - interpretatív beszéd egy előre megalkotott szöveg szó szerinti megszólaltatását jelenti, amikor a beszélő (színész, versmondó) azt szeretné elhitetni, mintha a szövegalkotás és elmondás szinkron tevékenység lenne, pedig a két művelet valójában nem egyidejű. A művészi interpretáció módja változott az évek során, hogy miért, erre igen találó kérdéssort állított össze Wacha Imre: a "ki, mikor, miért, hol s milyen eszközök által ad elő, illetve kinek s milyen körülmények között" kérdésekre adott válaszok adják a változás mibenlétét. Ez a kérdésbokor sokmindent rejt magában: a megszólaltatás helyét, a korstílust, az előadóművész egyéniségét, magát a verset, a "szűkebb és tágabb szövegkörnyezetet"stb... Ezen tényezők változása miatt kerülhet sor a versmondás történeti szempontú vizsgálatára. De vajon - ezeken kívül - mi teszi lehetővé az előadóművészek szabadságát az értelmezésben és az interpretálásban? Az írott szöveg korlátozott eszköztára, hisz az írott szöveg igen tökéletlenül jelzi a zenei elemek alkalmazását.

S most nézzük meg a három színművész versmondásában, hogyan változott az interpretatív stílus történeti megközelítésben! A három interpretációban a következő akusztikai elemeket vizsgáltam: a hangsúlyok száma és eloszlása, milyen lejtésformákkal élnek, mekkora e lejtésformák hangterjedelme, hangerőváltások, beszédtempó és szünethasználat. Kezdjük az akusztikai vizsgálatot a hangsúlyok számának és eloszlásának az elemzésével. A legtöbb nyomatékot Ódry Árpád alkalmazta, nála 264 a hangsúlyok száma, míg Latinovits 250, Mensáros 243 nyomatékot helyezett el a vers egészén. Önmagában ez az eredmény még nem sokat mond, hiszen nincs sok eltérés az egyes előadások között. Érdekesebb a kép, ha a hangsúlyok nyomatékfok szerinti megoszlását vizsgáljuk meg. Ez a következőképpen alakul: Ódry használta a legtöbb főhangsúlyt, számuk 19, Latinovitsnál négyet, Mensárosnál kettőt találtam. Ódrynál szinte minden szóra erős nyomaték esik. Véleményem szerint a három előadásmód közül az övé a leginkább érzelmi jellegű, ami azonban nem fedi azt, amit Molnár Gál Péter állít egy cikkében, hogy az Ódry-lemezek őrjöngő, búgó, éneklő színészt mutatnak. "Ódry előadóművészetét... a legforróbb vérmérséklet és a leghidegebb ész egyidejűsége, harca, szintézise jellemzi" (Csillag Ilona). Mensáros inkább közepes - és gyenge hangsúlyokkal él, ezáltal versmondása kevésbé lüktető, nagyobb szerepet kap az értelmi hangsúlyozás, az érzelmek csak árnyalják a gondolatiságot, meditációt.

Az egész versben mindössze két főhangsúlyt használ, ebből adódóan az összhatás visszafogottabb, mint Ódry előadása. Az elhallgatott költő-cigány önfelszólítása az írásra a legridegebb körülmények között is - mindez leginkább Latinovits előadásában szólal meg. Ugyancsak a hangsúlyeloszlás ad alkalmat arra, hogy megvizsgáljuk az egyes előadások érzelmi hullámzását, a vers melyik pontja emelkedik ki e szempontból a különböző interpretációkban. Ódry már a kezdettől erősebb hanggal és nyomatékokkal él, így nehezebb neki a vers hangulati ívét kidolgozni. nála a legerőteljesebb rész a 3. versszakra esik, ahol a 41 nyomaték a következőképpen oszlik meg: 2 fő, - 16 erős, - 15 szakasz, 8 gyenge hangsúly. Itt alkalmazza a legtöbb s eloszlásában a legerősebb nyomatékokat. Mensáros előadásában az érzelmi tetőfok a 6. és 7. versszakokra esik, ahol is a 6. szakasz 37 nyomatéka 9 erős, 22 szakasz és 6 gyenge hangsúlyban oszlik meg. Ebben a versszakban már emberiség méretűvé tágul a horizont, s az első két sorban megfogalmazott átokból egy hatalmas fohász nő ki "egy új világért". Hangulatilag ugyanezt folytatja a 7. versszak kezdő szava, majd ezt a fohászkodó hangot a " felismerés és tisztánlátás döbbenete vágja el".(Gáti József) Mensárosnál az utolsó sorok hangulatát a felhőtlen öröm, a teljes bizakodás jellemzi, ezt bizonyítja a versszak hangsúlyainak nyomatékfok szerinti megoszlása: a 32 hangsúly közül 2 a fő, 6 az erős, 13 a szakasz s 11 a gyenge hangsúlyok száma. Latinovitsnál a 6. versszak hangsúlyeloszlása megközelítőleg azonos Mensárosével (3 fő,- 3 erős, - 23 szakasz, 7 gyenge hangsúly), a 7. szakasz azonban másként alakul. Nála inkább a szakasz (16) és a gyenge (18) hangsúlyok dominálnak. Az ő hangjában nem jelenik meg az a felhőtlen öröm, amely Mensárosét jellemezte, Latinovits bizakodásában még benne van a félelem, az előző szakaszok átok- és fohászhangja. Annak vizsgálatából, hogy a hangsúlyok a mondat, tagmondat mely részére esnek, a következő eredményre jutottam: Ódry nagyon sok,szintaktikailag kevésbé fontos szóra is nyomatékot helyez, így a valóban fontos elemeket nehéz kihallani a versegészből; ezért is nehéz megvalósítania a vers hangulati ívét. Ezekkel a túlhangsúlyozásokkal előadásmódja deklamáló jellegűvé válik, azaz a hallgató számára így akarja érthetővé tenni a verset. Ezzel szemben a Latinovits és Mensáros előadásában ritkábban alkalmazott ilyen típusú nyomatékok sokkal "erőteljesebbek", kifejezőbbek, hiszen az a kevés számú fő- és erős hangsúly - amellyel dolgoznak - tartalmilag fontos helyre kerül: az egyre fokozottabb érzelmeket kifejező refrénekre és az érzelmi tetőpontot jelentő versszakokra.

Elemzésünkkel térjünk át az akusztikai elemek következő csoportjára, a hanglejtésre. A lejtésformák alakulásáról is készíthető hasonló összefoglaló "táblázat", mint a hangsúlyeloszlásról; éljünk itt most a szó szoros értelmében vett táblázattal:

Előadók / Hanglejtési formák	Ódry	Mensáros	Latinovits
emelkedő	21	1	4
ereszkedő	43	67	71
"lebegve előremutató"	25	22	15

Ezek az adatok is alátámasztják a hangsúlyeloszlás vizsgálatának eredményét; a legváltozatosabb lejtésformákkal Ódry él, fiatalabb társainál uralkodó az ereszkedő dallamforma. Ódry előadásmója a "leglátványosabb". Kihasználja a dallamformák változatosságát, ahogy kihasználta a hangsúlyok "erejét" is, hogy megértesse a verset a hallgatóval. Ezek a gyakori érzelemváltást tükröző dallamformák a megértetés eszközei, így jelenik meg itt is az érzelmi és az értelmi oldal kölcsönhatása. A Mensárosnál és Latinovitsnál többségben lévő ereszkedő és "lebegve előremutató" dallamformák is ugyanazt a szerepet töltik be, mint a közepes- illetve gyenge hangsúlyok, vagyis Ódryval ellentétben azt a benyomást keltik, nem lehet bizonyosság, hogy: " Lesz még egyszer ünnep a világon". Ódry előadásmódjának színességét fokozza az egy-egy tagmondat dallamvonalán belül található számtalan kicsúcsosodás. Ezek a dallamváltások adják előadásának lüktetését, ugyanakkor érződik belőlük a már említett megértetési szándék, amely megint csak előadásmódjának deklamativ jellegét erősíti.S ennél a pontnál át is ugrottunk a lejtésformák hangterjedelmének vizsgálatára. A legnagyobb hangmagasságbeli váltások - az érzelmi változásoknak megfelelően - Ódry interpretációjában fordulnak elő. Sokszor szintaktikailag kevésbé fontos helyen is nagy hangmagasságbeli váltással él, csakhogy a hallgató figyelmét állandóan ébrentartsa. Ezzel szemben Latinovits csak a valóban fontos szavaknál, gondolati csúcsoknál él ezzel az akusztikai lehetőséggel. A legkisebb hangmagasságbeli váltások Mensáros előadásában vannak. Ez megint csak meditatív jellegű előadásmódjával van kapcsolatban, hiszen a monologizáló stílustól távol áll az erőteljes akusztikai elemek megléte. Érdekes, hogy a gondolatiság hányféleképpen jelenik meg a három interpretációban: Ódrynál megértetési szándékú elemzési mód, Mensárosnál kezd belsővé válni, Latinovitsnál pedig már attitűdként jelentkezik. Ezt a különbséget fejezik ki az egyes szupraszegmentális elemek, pontosabban bennük realizálódik a változás. Kövessük nyomon a különbséget egy újabb akusztikai elem, a szünet segítségével! A szüneteloszlásról - mennyiségi és minőségi szempontból is - készítettem táblázatot.

Előadók / Szünetfajták	Ódry	Mensáros	Latinovits
rövid	85	66	56
közepes	6	23	26
hosszú	0	6	4

Ódry versmondásában egyetlen hosszú szünetet sem találunk, a legtöbbet Mensáros használja. Ennek magyarázata nagyon egyszerű mindannak ismeretében, amit eddig megállapítottunk. Az érzelmi színezetű versmondásban a különböző érzelmek egybesodorják a szólamokat, ezért nincs helye Ódrynál a hosszú szünetnek. Másik oka pedig az, hoyg a hosszú szünet mindig tartalmasabb, nagyobb jelentőségű szavak előtt jelentkezik; Ódrynál rengeteg ilyen szóval találkozunk, képzeljük el, mi lenne, ha betartaná ezt a szabályszerűséget.

Az olyan típusú interpretálásban, ahol kisebb szerepet játszik az érzelem, s az értelmi oldal a domináns, a lassúbb tempó hosszú szünettel párosul. Ez figyelhető meg Mensárod előadásában. Érdekesnek tűnik az a megfigyelés is, miszerint Ódry csak minden grammatikailag meghatározott helyen tart szünetet, ettől való - egyéni - eltérést nem találunk nála. Ezzel szemben Latinovits nem ragaszkodik ennyire az írott szöveg által nyújtott "szabályokhoz", sokszor egybemos tagmondatokat, vagy fordítva, elszakít grammatikailag szorosabban összetartozó részeket.
Ezek után térjünk át egy következő akusztikai jellemző vizsgálatára, amely összefüggésben áll a szünethasználattal, ez a beszédtempó. A beszédtempót úgy mértem, hogy megállapítottam, az első versszak első hangjától a utolsó versszak utolsó hangjáig mennyi idő telik el. A versegészre vonatkozó mérések eredménye így alakult: a legrövidebb idő alatt Ódry interpretálta a verset: 3' 17", a leghosszabb időre Mensárosnak volt szüksége: 3' 95", s a középértéket Latinovits képviseli: 3' 81". Hasonló eredményt mutatnak a tempóindexek is, ahol a számok a versegészre vonatkozó átlagot mutatják: a legmagasabb tempóindex Ódry előadására jellemző: 8,7 beszédhang/másodperc, Latinovitsé 6,5 , s végül Mensárosé 6,2. Természetesen a vers egyes szakaszain változik a tempóindex, hiszen befolyásolhatják a különböző érzelmek. Kísérletképpen megmértem, hogy az indító versszakhoz képest mekkora eltérés van az érzelmi tetőponton az egyes interpretációkban.

Ódry:	1. versszak: 8,9	--->	3. versszak: 9,07
Latinovits:	1. vsz:8	--->	6. vsz: 7,5
Mensáros:	1. vsz: 6,1	--->	6. vsz: 8,9

Ódrynál a fokozás minimális tempóbeli váltással jár, Mensárosnál fokozódik a beszédtempó, Latinovits pedig - éppen ellenkezőleg - laassítja a beszédtempóját a kiemelt versszakban. Ennek a tempólassulásnak is figyelemfelkeltő hatása van. A beszédtempó-tartományba tartoznak azok a ritmikai tényezők, melyek főképpen az egyén temperamentumától, lelkiállapotától függenek; vizsgáljuk meg, mennyiben tartják magukat az egyes előadók a vers által diktált ritmushoz, illetve mennyiben térnek el attól. A vers háromütemű tízes sorokból áll. Hármuk közül Ódry ragaszkodik leginkább ehhez a ritmushoz, nemigen találunk nála a ritmust felbontó szünetet, s ehhez a versritmushoz igazodik hangsúlyhasználata is, hiszen a legtöbbször az ütemkezdeteken alkalmaz erősebb hangsúlyt. Szünetei mindig a grammatikailag meghatározott helyeken találhatók. Mensáros előadói ritmusát inkább a gondolati, eszmei tartalmak határozzák meg, mint a vers grammatikai szabályai. Ő sokkal kevésbé ragaszkodik a nyelvi-, ritmikai előírásokhoz, s így gyakran hallani ilyen megoldásokat: " a zárát", " a zörvény". Latinovits azzal rúgja fel a versritmust, hogy nem használ semmiféle szünetet az ütemhatárok érzékeltetésére, egybemos tagmondatokat.
Ezekből az elemzésekből látszik, mennyire összefonódva jelennek meg a különböző akusztikai elemek. Elválaszthatatlan a többi zenei tényezőtől az utolsó akusztikai elem, amelyet megvizsgáltam: a hangerő és váltásai.
Az átlagos hangerő tekintetében a következő megállapítás tehető: Ódryé a vers egészén erős, Latinovitsé közepes, Mensárosé a leghalkabb. Ennél fontosabb azonban, hogy mennyire élnek a hangerő adta lehetőségekkel, az érzelmi váltásokat kísérik-e hangerőváltások. Ódry előadásában nincsenek nagy hangerőváltások, mivel kezdettől fogva közepesnél nagyobb az átlagos hangereje. Latinovits éppen a hangerő visszafogottságával, szinte suttogásszerűvételével emeli ki az érzelmileg, tartalmilag fontos részeket. Mensáros egy fokkal halkabb versmondásában is találunk hangerőváltásokat, hasonló helyeken, mint Latinovitsnál. A kettejük közötti különbséget az utolsó két

versszak hangerejének eltérő alakulása mutatja: Mensáros interpretációjában a hangerő egyre fokozódik a 6. versszakban, s a 7.-ben sem halkul el annyira, mint Latinovitsében. Ő jobban átadja magát az új világ eljövetele fölötti örömének, ezt mutatták más akusztikai elemek is.
Megvizsgáltunk tehát minden szupraszegmentális elemet a három előadóművész versmondásában, s ezzel három eltérő felfogást próbáltunk nyomon követni. Mindezen túl azonban szerettem volna bemutatni a versmondói stílus történeti alakulását, lehet-e valamilyen tendenciát megfigyelni az interpretatív stílus történeti vonatkozásában. Ehhez először is megpróbálom összefoglalni azokat a különböző művészi célokat, amelyek a versmondásokat alakították, s amelyek tükröződtek az egyes szupraszegmentális elemekben. Latinovits versmondása alkotó folyamat, az ő célja nem egyszerűen a megértetés, hanem a "gondolkodás sodrába való bevitel". Ő beemeli a hallgatót ebbe a folyamatba, őt is alkotóvá téve és kényszerítve. Számára az egész versmondás újrateremtése annak, amit előzőleg a költő világra hozott. Ezzel szemben Ódry interpretálása nem alkotói folyamat, hanem egy végtermék "átadása" a közönségnek. Az átadás mikéntjét már a versmondása előtt megalkotta. Az annyit hangoztatott ódry-i természetesség tehát viszonyítás kérdése: csupán az őt megelőző előadókhoz képest érvényes. Követőihez képest a deklamáló stílus képviselője. Sokkal nagyobb teret enged az érzelmeknek, mint "utódai". Nála a mindenki számára érthetővé tétel dominál, Mensárosnál a gondolatiság együttérzéssel párosul, Latinovits a gondolkodási folyamatba akar beemelni. Hadd nevezzem a versmondói stílus egyes fázisainak a következő lépéseket:

- a versmondás pillanatában az előadó már mint kívülálló jelenik meg + érzelmi oldal dominanciája (Ódry)
- belehelyezkedés a versegészbe az interpretáláskor (Mensáros)
- belehelyezkedés és mások bevonása a folyamatba + érzelmi dúsítás (Latinovits)
- belehelyezkedés és mások bevonása a folyamatba, de már nem annyira az érzelmekre, hanem az értelemre támaszkodva (például Jordán Tamás)

Mindezek igazolják azt a megfigyelést, hogy a versmondói stílus "alapköve" az érzelmi oldalról egyre inkább az értelmi oldalra tevődik át. Ez pedig nyelvi lenyomata annak az érzelmi intellektualizációs folyamatnak, amely korunkat egyre inkább meghatározza.

ZUR TYPOLOGIE DER INTONATIONSSTRUKTUREN (DIE THEME - RHEME - GLIEDERUNG)

Alla BAGMUT
Institut für Sprachwissenschaft
Akademie der Wissenschaften der Ukrainischen
SSR, Kiew, UdSSR

Zusammenfassung

Die akustischen Merkmale der Intonationsstruktur der verwandtschaftlichen slawischen Sprachen ermöglichen es, die typologische Charakteristik des Ausdrucks der aktuellen Gliederung der Aussage, sowie die überwiegende Anwendung der bestimmten Komponenten der Intonation in einer konkreten Sprache festzustellen. Der Ausdruck der aktuellen Gliederung durch akustische Mittel im Zusammenhand mit den Besonderheiten der Intonationsstruktur des Satzes als einer Redeeinheit und unter Berücksichtigung ihrer Semantik untersucht. Dabei gehen wir davon aus, daß beim Ausdruck der aktuellen Gliederung der Aussage die semantische Struktur des Satzes, seine kommunikative Aufgabe, die Wortfolge und die Intonation die dominierende Rolle spielen.

Einführung

Die Theorie der aktuellen Gliederung des Satzes, die von W. Matesius vorgeschlagen wurde, wurde in den Werken von vielen zeitgenössischen Linguisten weiterentwickelt. In der letzten Zeit wird die Aufmerksamkeit der Intonationstheorie immer mehr gewidmet. Die Frage zum Problem, wie die aktuelle Gliederung in der Intonationsstruktur der Aussage ausgedrückt wird, wurde auf der 11. Weltkongreß zu den Problem der phonetischen Wissenschaften erörtert (die Referate von Ch. Bonnot, I. Fougeron; R. Wenk; Č. Wan, Ch. Pan (I). Die Aufmerksamkeit der Linguisten ist dabei hauptsächlich den perzeptiven Charakteristiken und den Tonveränderung geschenkt. Wir halten es für notwendig, den Komplex der akustischen Mittel zu bestimmen

(F_0, Intensität, die Zeit der Lautung), die zum Ausdruck der aktuellen Gliederung dienen, unter Berücksichtigung der semantischen Besonderheit mündlicher Rede und der Intonationsstruktur der Aussage (2).

Methode

Die mündliche Rede (ukrainische, belorussische, tschechische, polnische, bulgarische Sprachen) wurde im Studio von 16 Muttersprachlern auf das Tonband aufgenommen. Der untersuchte Text wurde der auditiven und intonographischen Analyse unterworfen (Intonograph OPXAPC), die akustischen Angaben wurden statistisch bearbeitet. Der Untersuchung wurden zweigliedrige Aussagen zugrundegelegt, in denen der Bestand des Themas und des Rhemas auditiv festgestellt wurde.

Ergebnisse

Die akustischen Mittel, deren Funktion der Ausdruck der Aktualisierung der Aussage ist, sind durch die Struktur bedingt, sie sind Bauelemente der Satzintonation. So ist für die Erzählung die steigend-fallende (seltner -- die fallende) Tonbewegung, die fallende Intensität, die Verlängerung der Zeit der Lautung der Endsilben (der betonten sowie unbetonten). Diese allgemeinen Eigenschaften der Intonationsstruktur lassen sich als typologisch bedeutsam für die Intonation der Erzählung bezeichnen. Deshalb ist es nicht gerechtfertigt, in der steigenden (steigend-fallenden) Tonbewegung das Kennzeichen des Ausdrucks des Themas zu sehen, das sich gewöhnlich am Anfang des Satzes findet. Dieses Intonationsmerkmal ist durch die Intonationsstruktur des Satzes bedingt. Die Tonveränderung formt das Thema in einer bestimmten Satzposition, aber sie gestaltet auch das Rhema in derselben Position. Es ist auch falsch, die Verlangsamung der Lautung des Endvokals (oder der Endsilbe) als Unterscheidungsmerkmal des Rhemas zu betrachten, das gewöhnlich in der Endposition des Satzes liegt und automatisch betont ist. Demgemäß verlangt die Analyse

der Intonation der aktuellen Satzgliederung die Untersuchung des akustischen Ausdrucks des Bestandes des Themas und Rhemas und ihrer Unterscheidungsmerkmale.

In der Intonationsstruktur der Aussage sind die Veränderungen in der Anfangs-, Mittel- und Endposition zu unterscheiden. Die Intonationsveränderungen des Themas und Rhemas werden im Zusammenhang mit diesen Positionen bestimmt. Dabei stellt es sich heraus, daß die gewöhnliche Endposition des Rhemas und die Anfangsposition des Themas (die Grenze zwischen Thema und Thema teilt auch die Mittelposition ein) mit Hauptmerkmalen der Intonationsstruktur verbunden sind. Die steigend-fallende Toncontoure des Aussagesatzes bedingt oft die maximale Größe der Frequenz in der Anfangsposition. Dasselbe kann man auch in der Intensitätkurve beobachten. Gleichzeitig zeigt uns die auditive Wahrnehmung die größere Tonstärke des Rhemas als des Themas. Daraus erfolgt, daß die kommunikativ-semantischen Verhältnisse, die in der aktuellen Gliederung ausgedrückt werden, mit der akustischen Struktur der Aussage nicht übereinstimmen.

Die Analyse der akustischen Angaben zeugt davon, daß die maximale Größe Fo sowohl in der Anfangs=als auch in der Mittelposition liegen oder auch einige Positionen umfassen kann. Aus der Analyse der allmöglichen Varianten der Thema-Rhema-Positionen in der Aussage erfolgt, daß die maximale Größe Fo in meisten Fällen dem Thema gehören (59% aller Sätze) und viel seltene - dem Rhema (29%); in anderen Sätzen (12%) verteilt sie sich über das Rhema und Thema der Aussage. In einer jeden untersuchten Sprache ist diese Größe nicht konstant (Schema I).

Funktionale Charakteristik der Abgrenzung des Themas und Rhemas nach der maximalen Größe der Frequenz in Prozent

Schema I

Sprache	Thema	Rhema	Thema u. Rhema
Ukrainisch	61	21	18
Belorussisch	45	49	6
Polnisch	60	27	13
Tschechisch	75	14	11
Bulgarisch	54	34	12

Die detaillierte Charakteristik der maximalen Größe Fo nach Satzpositionen kann folgenderweise dargestellt werden (Schema 2):

Position der maximalen Größe Fo in der Satzstruktur in Prozent — Schema 2

Sprache	Satzposition			
	Anfangs-	Mittel-	End-	in Thema u. Rhema verteilt
Ukrainisch	47	15	3	35
Belorussisch	37	24	15	24
Polnisch	61	10	5	24
Tschechisch	61	8	3	28
Bulgarisch	60	21	-	19

Die prozentuelle Charakteristik der maximalen Größe Fo in der Struktur der Satzintonation in den slawischen Sprachen sieht folgenderweise aus: in der Anfangs = (53%), in der Mittel=-(15%), in der Endposition - (6%); in allen Positionen verteilt - 26 %.

Die Thema-Rhema-Gliederung wird durch das maximale Frequenzintervall ausgedrückt. Es umfaßt: im Rhema - 60%, im Thema - 35 %, im Rhema u. Thema verteilt - 5 %.

Das maximale Frequenzintervall findet sich in den slawischen Sprachen am häufigsten in dem Satzabschluß (36 %); in der Afangs- (26%) und Mittelposition (25%) ist es gleich; in anderen Fällen (13%) ist es im Satz verteilt. In der polnischen Sprache (47%) kommt das maximale Frequenzintervall im Satzabschluß am konsequentesten vor.

Wir betrachten auch die Intensitätsveränderungen im Satz: in allen slawischen Sprachen ist maximale Intensität für das Thema charakteristisch (64--86%).

Wir haben die durchschnittliche Zeit der Silbenlautung des Themas (T_1) und Rhemas (T_2) festgestellt. Ihr Verhältnis (T_2:T_1) zeigt die Stufe der Versammlung oder Beschleunigung der Lautung des Rhemas (im Vergleich zu der des Themas).

Die Abhängigkeit $T_2:T_1$ ist sowohl durch formale als auch durch semantisch-kommunikative Faktoren bedingt. Dieses Verhältnis ist für die untersuchten Sprachen gleich 1,276, d.h. das Rhema wird etwas langsamer als das Thema ausgesprochen. Die maximalen Tempounterschiede im Rhema und Thema wurden in der ukrainischen Sprache festgestellt (1,480), die minimalen -- in der belorussischen (1,143).

Typologisch beceutsam für den Ausdruk der Thema-Rhema-Gliederung sind folgende akustische Charakteristiken:

- die Verlangsamung bei der Lautung des Rhemas (1,3 mal);
- die maximale Frequenzintervall im Rhema;
- die maximale Größe der Tonfrequenz im Thema;
- die maximale Intensität im Thema.

Literatur

1. BONNOT, Ch. --Fougeron, I.: Intonation et thematisation en Russe modern (Proceedings XIth ICPhS. - Tallinn, 1987.Vol. 2. P. 463--467.
2. BAGMUT, A. I.: Struktura i funkcional'no-semantičeskij aspekt intonacii prostogo povestvovatel'nogo predloženija v slavjanskih jazykah. Moskva 1980.

PROSODY OF CONFERENCE SPEECH IN ENGLISH, HUNGARIAN AND RUSSIAN

József BENDIK
National Commission for UNESCO
Budapest, Hungary

The present paper covers part of a cross-linguistic investigation of the prosody of conference speech in English, Hungarian and Russian.

Speech prosody is a collective term for variations in pitch, loudness, duration, pause and voice quality. Stress, tones, juncture, rhythm and timbre are prosodic effects meaningful at syllabic, syllable sequence and utterance levels. Sense distinction, speech flow segmentation, prominence, indication fo grammatic relations, of attitudes and emotions are functions of speech prosody (Cf. ref. 2, 3, 4).

For the purposes of this study a functional speech variety was defined as oral speech the prosodic features of which are delimited and affected by identical sets of extralinguistic factors. Conference speech (CS) is a functional speech variety the prosodic characteristics of which are delimited by the medium of the language symbol, the number of speakers, presence of audience, purpose of the speech act, degree of its preparedness and intention of the speaker. Accordingly, CS is labeled as reproductive, public conference statement (Cf. ref. 5). In order to reveal the characteristic features of CS it was opposed to the functional speech variety of reading aloud (RA) described in terms of delimiting labels as reproductive, non-public, unprepared. When establishing the corpus of recordings an attempt was made to achieve that the prosodic features of both CS and RA be equally affected by factors related to the person of the speaker, such as: sex, age, social and educational background, speech proficiency, idiosyncracy.

The aim of the study was to offer to teachers and users a language-independent method and system for description of prosodic features which would allow them to identify the similarities and differences between the three languages and to devise appropriate teaching and learning activities. Some of the data obtained by auditory, graphical and numerical processing of phonetic realizations could be of phonological importance, e.g. in the search for language universalia.

The corpus was made up of utterances picked from recordings of statements by English and Russian native speakers at UNESCO conferences so as to represent the following frequently occurring meanings and structures in CS: opening, key and concluding utterances of the text; attitudinal-emotive meanings such as appeal, approval, objection, anxiety; syntactic relations of enumeration, parenthesis, clause coordination and subordination. The selected utterances were translated into the other language and then into Hungarian. Three experienced conference speakers and natives of the three languages were instructed to record the selected utterances twice: firstly, to correspond to the definition of RA and secondly, to that of CS.

The graphical metalanguage (Fig. 2) was designed to mark for every syllable the pitch height, loudness and duration with a single note. At the inter-syllabic level the tones, intervals (not to be confused with intervals of the musical scale) and mode (glide and non-glide) of pitch movement were marked. This allowed a detailed depiction of the features perceived as well as a numerical expression of such data as the mean pitch

height of tone units and utterances, mean syllabic duration, per cent of sustained syllabic pitch, per cent of glides, size of inter-syllabic intervals and pitch ranges.

A tone unit was defined as a meaningful stretch of the speech flow flanked by pauses. Because in the three languages differing kinds of pause were perceived only absence of phonation coupled with a shift in pitch was considered as a boundary pause. The traditional British tone unit structure of prehead, head, body, nucleus, tail which allowed for one or two prominent syllables had to be extended to accomodate more peaks of prominence within the body. The nucleus was defined as the last tone-bearing prominent syllable of the unit.

The functions of prominence and expression of attitudes were found to be the most characteristic of CS in English, Hungarian and Russian.

A common feature of prominence in the three languages is that it is achieved by means of a combination of at least three variables: pitch, loudness, and duration. Relative pitch height and pitch movement were the easiest to perceive by ear and to identify with meaning. To differing extents in the three languages the following features of pitch movement were found to be meaningful: tones (sustained, falling, rising and complex), mode (glide or non-glide), interval (between two adjacent syllables) and range (at syllable sequence levels). The average pitch height of all the examined utterances of CS is the highest in English and the lowest in Hungarian. This is attributed to the fact that in English prominence is generally achieved by means of bigger intervals and wider ranges of pitch movement. The lowest mean pitch height of Hungarian could be explained by narrower ranges and intervals as well as by the fact that word stress is fixed on the first syllable which reduces pitch movement within words. A characteristic means of prominence in Hungarian CS is the emphatic juncture followed by a rise before and a fall after the nucleus with the rest of the tone unit sustained near the base pitch level (Fig. 1, Line 2).

While in English and Hungarian prominence is generally achieved by a rising of the pitch before the prominent syllable, in Russian downward prominence was very frequent, especially in utterances conveying attitudinal meanings. The prehead, head and body of such tone units are sustained at high ptich levels and followed by a fall before the nucleus. The tail would then be sustained at base pitch level (Fig. 3, Line 10).

Closely connected to the function of prominence is that of conveying attitudes and intentions of the speaker by means of tones. Commonly for the three languages in RA all tones ended in a rise to express incompleteness or in a fall to express completeness. Sustained endings were not perceived at all (Cf. ref. 1). In CS rising tones meant to signal incompleteness were substituted for falling or complex ones, i.e. the function of attitudes dominated over that of segmentation of the speech flow or indication of grammatic relations (Fig. 1, Lines 1-2, 5-6, 9-10).

On the basis of numerical data obtained it was concluded that glides and complex tones of single syllables were typical of CS in English and did not occur either in Hungarian or in Russian. In Hungarian the number of complex tones was the lowest. In Russian slight differences between tones preceding the nucleus were found to be a means of expressing subtle differences of speaker attitude. In Figure 4 the first variant of the utterance of address was said by native speakers to sound "formal", the second: "polite", the third: "ironical".

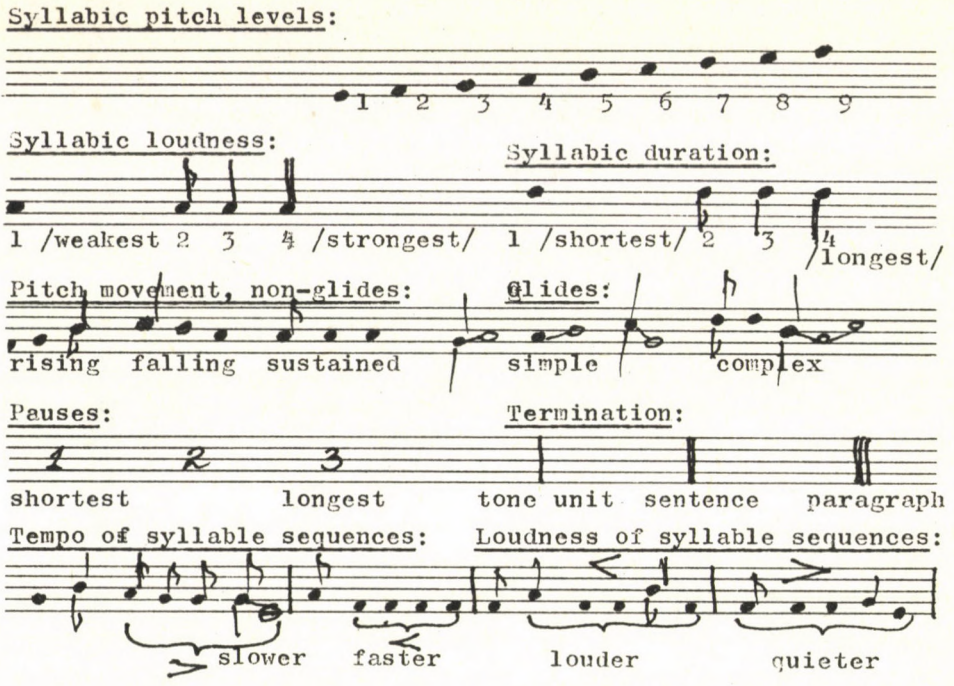


Figure 2. The metalanguage

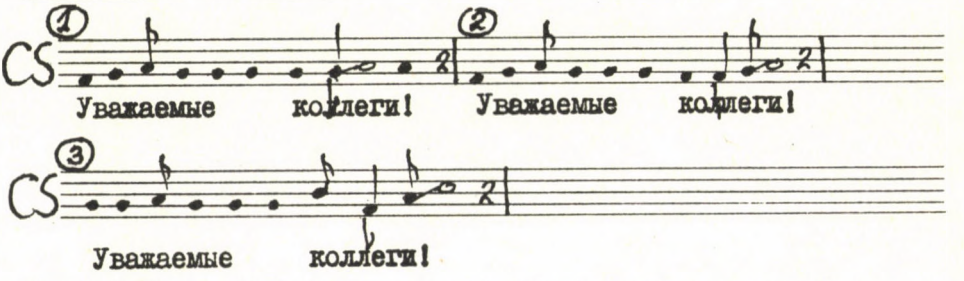


Figure 3. Three attitudinal meanings in Russian conference speech

References:
1. BRAZIL, D., The Communicative Value of Intonation in English, Birmingham, English Language Research and Bleak House Book, 1985.
2. BRIZGUNOVA, E.A., Zvuki i intonacija russkoi rechi, Moskva, izdatel'stvo "Russkii jazik", 1977.
3. CRYSTAL, D., Prosodic Systems and Intonation in English, Cambridge, Cambridge University Press, 1969.
4. VARGA, L., "Prozodémák a magyar beszédben és jelölésük az intonációs átiratban" in: "Műhelymunkák a nyelvészet és társtudományai köréből", Budapest, MTA Nyelvtudományi Intézete, 1987.
5. WACHA, I., "Az elhangzó beszéd főbb akusztikus stíluskategóriáiról" in: Általános Nyelvészeti Tanulmányok X, Budapest, 1974.

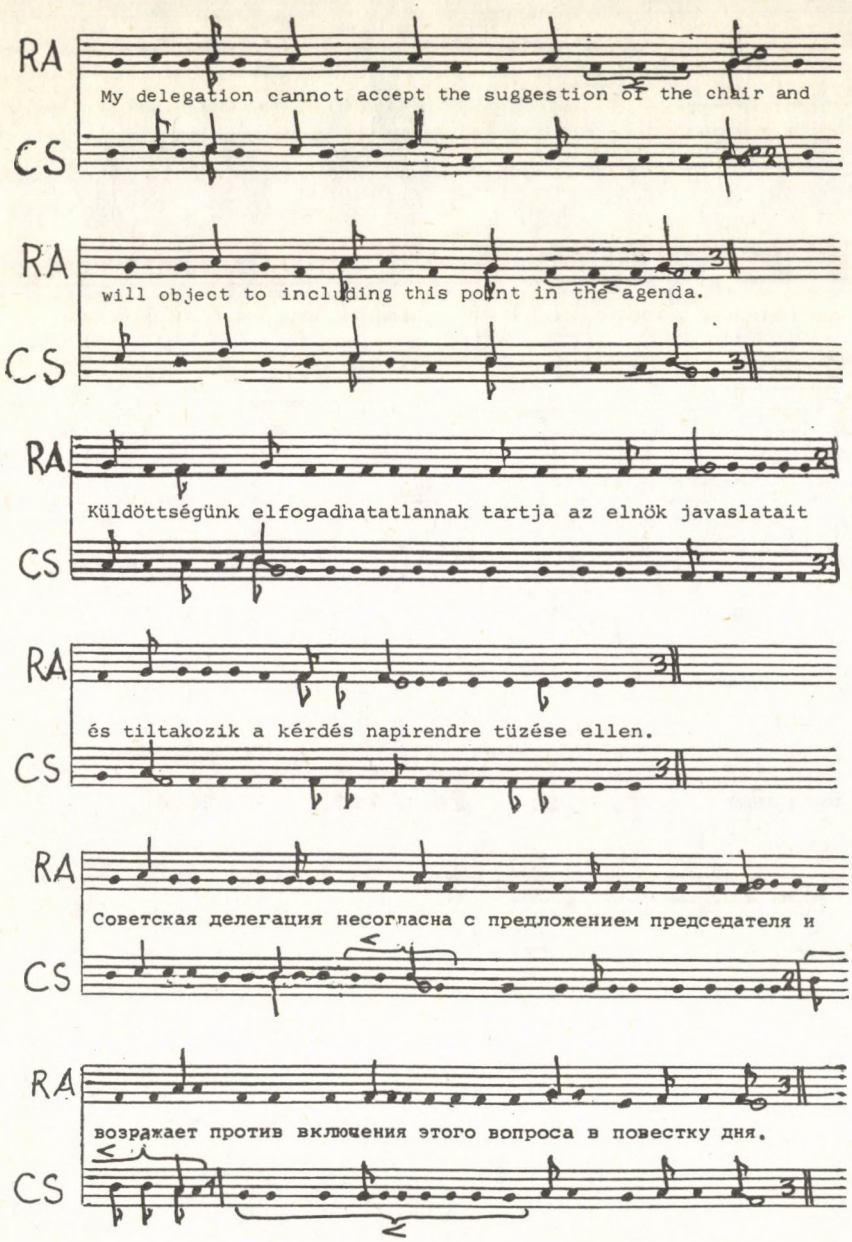


Figure 1. An utterance conveying the attitude of objection.
RA stands for reading aloud and CS for conference speech

GEFÜHLSVERARMUNG IN DER GEGENWART; WEITERES ZUR PROBLEMATIK

Béla BÜKY
Institut für Linguistik der Ungarischen
Akademie der Wiss., Budapest, Ungarn

Als Fragestellung

Nach der Meinung von András Vértes O. (1, 12--18 und 300) zeigt die Sprache, ja das ganze Kommunikationsvermögen von heute überall -- so auch in Ungarn -- klare Zeichen von Gefühlsverarmung. Die Zeichen hierbei z.B.: negative Änderung des Mutter-Kind-Verhältnisses, allgemeine Verachtung des Gemüts, Überwerten der Rolle des Intellekts, Reizüberangebot, Abnehmen in Kontaktfähigkeit, sog. Verkopfungsneurose, Mode der Gefühllosigkeit im Stil, in den Ausdrucksformen im allgemeinen. Die sprachliche Zeichen betreffend bezieht sich Vértes teils auf phonetische (1, 20--29) wie auch lexikalische (1, 151--154), teils auch auf stilistische bzw. sog. extralinguale (1, 155--162 bzw. 288--292) Erscheinungen.

Einige diesbezügliche Bemerkungen von Vértes über phonetischem Feld:

Vértes bezieht sich auf Béla Zolnai, der behauptet, daß "Je gefühlvoller eine Sprechäußerung ist, desto melodischer ist ihr Verlauf" (1, 20). Die Melodielosigkeit der Sprechtätigkeit im allgemeinen wäre also ein Zeichen für die Gefühllosigkeit von heute.

Vértes bezieht sich auch darauf, daß die Rhythmik der Sprechtätigkeit unserer Zeit -- verglichen mit der z.B. zur Jahrhundertwende -- sich geändert hat, und zwar in die Richtung von Arhytmik. Als Grund dafür erwähnt Vértes diesbetreffend den rhytmusnivellierenden Einfluß der Schriftsprache.

Vértes erwähnt weitere phonetische Merkmale. Er erwähnt z.B. Fónagys Meinung (2, 226--227) wonach die Sprechsätze in unseren Tagen eher stoßartig wirken, während die Sätze zur Zeit von Jahrhundertwende ein mehr oder weniger melodische Spannung aufweisen (Vértes über Fónagys Meinung: 1, 29).

Einige Bemerkungen von Vértes über lexikalischem Feld:

Nach phonetischen Erscheinungen führt Vértes ziemlich viele lexikalisch-lexikologische Belege zur Beweisführung seiner These an: Abnahme der Diminutivformen, der Frequentativa, ja Abnahme der Bildungssuffixe im allgemeinen usw.

Die Stilistik und Extralinguistik betreffend erwähnt Vértes die Abnahme der bildlichen Ausdrücke, der Metaphern, der Metonymien usw. Schließlich unternimmt Vértes eine Analyse der extralingualen, also der außersprachlichen Effekte (Lachen, Weinen, Gebärden und Gesten) und hier sieht er auch feste Beweise für die Abnahme der letzteren.

Unsere Thesen

Die Richtigkeit der obigen Behauptungen von Vértes will ich im Grunde anerkennen. Allerdings möchte ich darauf aufmerksam machen, daß der Mangel von positiven, humanistischen, brüderlich heilsamen Gefühlen gar nicht dem Mangel der Gefühle im allgemeinen (These Nr. 1), und daß der Mangel von Ausdruckserscheinungen der Gefühle gar nicht dem Mangel der Gefühle (These Nr. 2) gleichzusetzen sind.

Einige Bemerkungen zur These Nr. 1

Die Zunahme der gefühlsintensiven Manifestationen (obwohl mit negati-

vem Inhalt) zeigt sich z.B. in einem häufigeren Gebrauch von Grobheiten, Unflätigkeiten, Flüchen, Fluchworten sowie von den begleitenden oder ersetzenden Gebärden, Zeichen, mimischen Ausdrucksformen. Der Gebrauch eines verstärkten Sprechtons, einer speziellen Sprechmelodie, (Überintonation), Abnahme der sog. sprecherischer Disziplin usw. gesellen natürlich zu den obigen Erscheinung lexikalischer Art. (3)

Einige Bemerkungen zur These Nr. 2

Diese These möchte ich eingehender als die vorherige erörtern. Ich sehe diese Frage so, da3 wir -- speziell in Ungarn -- auf zwei wichtige Gründe hinweisen sollen, die den Mangel der Ausdruckserscheinungen von Gefühlen hervorgerufen haben bzw. haben konnten:

1. Angst, Furcht, Unterdrückung der aufrichtigen Gefühle. In einem politischen System, worin wir in den drei vorangehenden Jahrzehnten -- sagen wir zwischen 1957--1986 -- gelebt haben (und das war die Situation für viele Ungarn von Nachbarländern) galt vor allem der Grundsatz: "nichts sagen, kein Kopfweh haben" und auch die Kinder, ja selbst in Vorschulalter, waren von den Eltern dazu erzogen. Das daraus erwachsene bewu3t blasierte, gefühllose Verhalten ist auch die eine Ursache der scheinbaren Gefühlsverarmung (meiner Meinung nach).

2. Ein anderer Faktor der Gefühlsverarmung liegt in einem gewissen Schamgefühl bzw. Selbstbewu3tseingefühl.

Imre Wacha erzählte einmal (im Privatgespräch), da3 er jemanden telephonisch angerufen hatten, worauf der betreffende -- nach einigen Worten, aus unbekannten Gründen -- den Hörer auf den Tisch gelegt hat; dabei konnte man leise Stimme eines Zwiegesprächs (mit Lachen, mit Lautstärkeschwankungen, usw.) mithören, welches sich nach den fürs Radio typischen textphonetischen Merkmalen (etwas verschleierte, weniger konturierte Stimme usw.) als eine Rundfunksendung erwies. Und noch etwas: eine gekünstelte Natur der au3ersprachlichen Effekte (Lachen, Häsitation, moroser Sprechton, Intonation) sprach auch für Radiosendung. Durch Einschaltung eines Radioapparats konnte Wacha seine Annahme bestätigen. Nun, derartige unnatürliche lautliche Ausdrucksformen der Gefühle werden im Alltag gewöhnlich vermieden. Es ist ein heilsamer Vorgang im Gefühlleben von heute, da3 man solche, nicht naturgemä3e emotionelle Klischees widerlich findet. Und die Folgeerscheinung: der Alltagsmensch zieht sich -- anstatt seine Gefühle klar zu äu3ern -- in sein "Schneckenhaus" zurück. Diese Form der sog. Gefühlsverarmung hat also fernsehspezifische, massenkommunikative--antimassenkommunikative Gründe.(4)

Weiteres für die zwei Thesen

Die zwei erwähnten Gründe für die Gefühlsverarmung (These 1 und 2) hängen eigentlich zusammen: die strenge staatliche Übersicht der Gewissensfreiheit, die als einer Grund die Gefühlsverarmung (These 1) hervorruft, beeinflu3t auch die Massenkommunikation auf solche Weise, da3 es dadurch am allerm eisten nur typisch rundfunk- bzw. fernsehspezifische (also kontrollierte) Gefühlsäu3erungen übertragen werden.

Einige Bemerkungen zur These 2 (Punkt 1)

Es ist aber unverkennbar, da3 gerade in unseren Tagen diese Verdrängung der Gefühlsäu3erungen (das Zurückgesetztheitsgefühl, Minderwertigkeitskomplexe bei einigen Schülern -- politische Selbstzensur und Zurückhaltung

bei Erwachsenen) zu lockern beginnt: solche Beispiele lieferte z.B. der etwas wehmütiger Unterton in der Rücktrittserklärung der Parlamentsabgeordnete Frau Cservenka (in Morgenchronik von Radio Kossuth, den 3. März 1989) sowie der begeisterte Ton der Redner von dem Nationalfest am 15. März in Ungarn.

Was den Punkt 2 der These 2 betrifft, hier sehe ich noch keine Änderung, kein Lockerwerden der Gefühlsverarmung, der Entfremdung im Gefühlleben (5).

LITERATUR

1. VÉRTES O. András: Érzelmi világunk és a nyelv történeti változásának kölcsönhatása I-III. Magyar Nyelv 82 (1986), 11--29, 151--162, 288--300. -- Auch als ein Band (Nr. 179) des Serienwerkes Magyar Nyelvtudományi Társaság Kiadványai (mit einigen Ergänzungen). Budapest, 1987. Magyar Nyelvtudományi Társaság.
2. FÓNAGY Iván: Une histoire contemporaine: changements linguistiques en Hongrois dans la période 1900--1940. Etudes Finno-ougriennes 10. 1973, 215--230.
3. VÉRTES O. András: Language usage and the growing frequency of emotion with tension and excitement. Magyar Fonetikai Füzetek 19. 1988, 92--106.
4. WACHA Imre: Beszéd: szituáció, szöveg és hangzás együttese a rádióban és a televízióban. In: Grétsy László: Nyelvészet és tömegkommunikáció. Bd. 1. S. 7--246. Budapest, 1985. Tömegkommunikációs Kutatóközpont.
5. HELLER Ágnes: Az ösztönök. Az érzelmek elmélete. Budapest, 1978.

ON THE NATURE OF FOREIGN ACCENTS

Una CUNNINGHAM-ANDERSSON and Olle ENGSTRAND
Department of Linguistics
Stockholm University, Stockholm, Sweden

Introduction

This paper reports work carried out as part of a research project on native Swedish speakers' attitudes to foreign accent. One part of the project is to use native listeners' reactions to accented speech samples as an indirect way of evaluating hypotheses on attitudes that are basically conditioned by geographical, cultural, or ethnical factors (1). The other part of the project takes a linguistic-phonetic view, its purpose being to form and evaluate hypotheses concerning native Swedish speakers' attitudes to various phonetic characteristics of foreign accent. It is an interesting observation, for example, that certain phonetic features, existing in parallel in different languages, may have quite different functions in those languages. When studying the phonetics of the Lappish language, for example, we have noted a strong utterance-final F0 fall coupled with a prominent, subsequent aspirative noise (2). Whereas these effects are apparently perfectly regular syntactic markers in Lappish, they are likely to be perceived as an emotional expression (of resignation or the like) when heard by a Swedish listener. Since native pronunciation habits are known to be carried over to the speaker's second language, such phonetic-semantic clashes may well corroborate already existing, positive or negative beliefs about certain groups of non-native speakers.

Before we can investigate native Swedish speakers' attitudes to phonetic characteristics of foreign accent we must know more about these characteristics. At least two reasonable hypotheses can be advanced in response to the question of what makes a particular accent sound foreign. Firstly, the impression of foreignness as judged from listener reactions increases as a function of the number of phonetic deviations in a speech sample (Hypothesis 1). In contrast to this straightforward, quantitative hypothesis, a different, qualitative assumption would be that native listeners are not primarily sensitive to the amount but rather to the kind of deviation. It could be hypothesized, in other words, that some deviant characteristics have a strong tendency to create an impression of foreign accent, while others might sound like a possible (though not necessarily existing) regional accent, or merely peculiar (Hypothesis 2). If this is true, we need to find out which those respective features are. Accent strength is related to this question. Our hypotheses here are that the perceived strength of a foreign accent increases as a function of the number of phonetic deviations present in the speech sample (Hypothesis 3) but that some combinations of deviant characteristics give an impression of stronger foreign accent than others with the same number of deviations (Hypothesis 4).

The first experiment (Experiment 1) was designed as a preliminary test of these hypotheses to obtain an indication of the phonetic conditions required to give the impression of foreign accent. An additional purpose of Experiment 1 was to explore the possibility that particular sets of deviant phonetic characteristics can be used as "signatures" in identifying a particular foreign accent. Consider, for example, a "Finnish" accent of Swedish. Current work on analysis of genuine Finnish accents suggests that typical features may be lack

of an F0 correlate of the Swedish grave accent, unaspirated initial plosives, exaggerated length contrasts in consonants and vowels, velarized /l/, and alveolar rolled /r/. It may be the case that certain combinations of these characteristics occurring in otherwise native Swedish tend to give an impression of a Finnish accent whereas other combinations, or the presence of one single characteristic, would not give such an impression. Our hypothesis is that these five features constitute relatively strong predictors for judging a given speech sample as originating from a native speaker of Finnish. (Hypothesis 5).

Experiment 1

After careful training, a phonetician (the second author) recorded a large number of readings of a version of "The North Wind and the Sun" in Swedish. They differed in that the reader introduced one or more deviation from his normal native Swedish pronunciation into each reading. The speaker's accuracy and consistency in introducing these features into his speech were checked by measurements made from computer-generated spectrograms, LPC-spectra and dB-expanded oscillograms. The deviations are some of those which commonly occur in immigrant Swedish, with particular attention paid to characteristics of Finnish accents in Swedish:

1 The Swedish grave word accent is replaced by the acute accent.
2 The acute accent is replaced by the grave accent.
3 The speaker's usual tongue blade /r/ is replaced by its uvular equivalent.
4 The Swedish supradental (apico-post-alveolar) consonants are replaced by the corresponding dental consonants (/n,l,s,t,d/) + /r/.
5 Initial voiceless plosives are pronounced without their usual aspiration.
6 The usual durational distinction between long and short vowels and consonants is not made.
7 Vowels in unstressed syllables are reduced.
8 Post-vocalic /r/ is omitted.
9 Post-vocalic /r/ is replaced by lengthening of the vowel, which takes the quality of the nearest British English vowel.
10 /l/ is velarized.
11 Quantity distinctions in vowels and consonants are exaggerated.
12 Vowels are replaced as shown: /ø/-/o/, /ʉ/-/u/, /y/-/i/.
13 Vowels are replaced as shown: /u/-/o/.
14 /r/ becomes an alveolar roll.

34 of these readings were selected for use in the listening experiment, representing 14 readings containing single, artificially introduced deviant pronunciations, 10 readings with combinations of two deviant features, 5 readings with three deviant features, 3 readings with four deviant features, and 2 readings with five deviant features in combination. An unanalyzed imitation of a Finnish stereotype was also included for comparison. The 35 readings were played to 35 monolingual Swedish secondary school students (16-18 years old) in random order. The listeners were asked to indicate on a special form whether they perceived each of the readings to sound like (a) a foreign accent, (b) a possible regional accent of Swedish or (c) merely strange. If they were reminded of a particular accent they were to name it. They were also required to grade the reading for degree of deviation on a five-point scale, where 0 corresponded to "no deviation" and 4 to "maximum deviation".

Table I lists the deviations (see above list) which are involved in

each reading, the mean degree of deviation assigned by the naive informants to each reading, as well as which readings at least 50% of the informants perceived as foreign (F) or regional Swedish (S) accents or as merely peculiar (P). The fact that several readings were indeed judged as being foreign, several as regional accents of Swedish and several as peculiar sounding provides us with corroboration for Hypothesis 2: it is possible to introduce deviant phonetic characteristics into a reading such that listeners perceive foreign or regional accents.

Table I: Naive judgements of accentedness and foreignness

deviations	Mean Grade	Judged by >50%	deviations	Mean Grade	Judged by >50%
6	.64	S	1,14	1.76	-
13	.74	S	12	1.79	-
5	.97	F	1,7	1.82	-
1	1.06	S	12,13	1.86	-
2	1.09	S	1,5,12,13	2.00	F
5	1.12	-	1,2,5	2.03	F
1,6	1.26	P	1,5,10,14	2.09	F(Finnish)
3	1.29	S	1,12,13	2.09	S
4	1.31	-	12,14	2.12	F
8	1.45	S	1,5,10	2.21	F(Finnish)
7	1.47	-	1,6,12,13	2.21	P
1,2	1.49	F	1,12,14	2.24	F
1,5	1.57	F	7,9	2.44	-
14	1.62	-	7,9,13	2.45	P
1,14	1.71	-	9	2.71	P
10	1.74	F	1,5,12,13,14	2.77	F
1,10	1.74	F	1,5,10,11,14	3.32	F(Finnish)
			unanalyzed	2.94	F(Finnish)

The minimum needed to create the impression of a foreign accent (as judged by at least 50% of the listeners) is the presence of one of the following four features or combination of features: a)the initial voiceless plosives are pronounced without aspiration (feature no. 5); b) /l/ is velarized (no. 10); c) the vowels /y ø ʉ/ are replaced by /i o u/ respectively and /r/ is an alveolar roll (nos. 12, 14); d) the Swedish grave accent is replaced by an acute accent and the acute accent is replaced by a grave accent where this is possible (nos. 1, 2). All other combinations of features which were judged as sounding foreign by at least half the informants contained one of these four minimal possibilities.

Statistically significant positive correlations were found between the number of listeners perceiving a given reading as sounding like a foreign accent and the number of deviant features in the reading ($n=34$, $r=0.53$, $p(r)<0.01$) (which corroborates Hypothesis 1, that the impression of foreignness increases as a function of the number of deviant features in the reading); and also between the perceived strength of the foreign accent (as reflected in the average degree of deviation assigned to the readings by the listeners) and the number of deviant features in the reading ($n=34$, $r=0.74$, $p(r)<0.01$) (which corroborates Hypothesis 3, that the impression of accent strength increases as a function of the number of deviant features in the reading). Hypothesis 4, that some combinations of deviant characteristics give an impression of stronger accent than others with the same number of deviations, is clearly corroborated by the data summarized in Table I. Readings with single deviant features, which sounded

foreign to at least half of the listeners, were assigned average grades of deviations of between 0.97 and 1.74. Hypothesis 5 is corroborated by the fact that at least 50% of the listeners indicated that four readings (3 respectively containing three, four and all five of the above-named phonetic characteristics found in the speech of some native Finnish speakers as well as the unanalyzed imitation of a Finnish stereotype) were reminiscent of Finnish accents of Swedish.

It may be the case that other combinations of two or three deviations than those used above would also suggest a Finnish accent to most of the informants. Experiment 2, was designed to examine the components of a Finnish accent in Swedish more closely, and test the hypothesis that several different combinations of deviant phonetic characteristics can induce the impression of Finnish accent and that no particular characteristic must be present (Hypothesis 6).

Experiment 2

The method for this experiment was similar to that used for the previous one. The text was again read by the second author, both normally and with all possible combinations of one or more of deviations 1 (grave accent is pronounced like acute accent), 5 (deaspiration of plosives), 10 (velarized /l/), 11 (exaggerated quantity distinctions in V and C) and 14 (/r/ is an alveolar trill). These features can all occur to varying extents in the Swedish spoken by native Finnish speakers, although they also occur in other foreign accents of Swedish. This gave 32 versions of the text which were arranged in random order and played to a new group of 39 secondary school students. The listeners were required to indicate on a special answer sheet (a) whether they thought each reading sounded like a native speaker of Finnish speaking Swędish and (b) how strong they judged the reader's foreign accent to be (on a scale from 0-4).

Apart from one pair of features, three of the five deviant characteristics (in any combination) were required for a reading to be perceived as a Finnish accent by at least half of the informants. Beyond this minimum, the addition of more deviant features increases both the number of listeners hearing the reading as a Finnish accent ($r=0.91$), and the estimated accentedness of the readings ($r=0.92$).
Hypothesis 6 (that different combinations of features can give an impression of a Finnish accent in Swedish) is therefore corroborated.

The results of Experiments 1 and 2 suggest that it is possible to artificially build up the impression of a foreign accent by combining single phonetic deviations. In these experiments, artificial accents have been performed using a natural voice. The next step will be to replicate parts of these experiments by manipulating naturally-produced speech using an LPC-based synthesis technique.

References

1. CUNNINGHAM-ANDERSSON, U., ENGSTRAND, O.: Attitudes to immigrant Swedish - a literature review and preparatory experiments. Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS) 8. 1988, 103-152.

2 ENGSTRAND, O: Preaspiration and the voicing contrast in Lule Sami. Phonetica 44. 1987, 103-116.

ACCENT ET TON EN VIETNAMIEN

ĐỖ THẾ DŨNG
Université de Ho Chi Minh ville/Université Paris III et VII

I. CORPUS

Le corpus analysé se compose de 58 mots composés, dont 31 dissyllabes, 19 trisyllabes et 8 mots de plus de 3 syllabes).

Ces mots sont enchassés dans une phrase cadre - "Từ ___ được đọc ba lần"/ le mot ___ est lu trois fois - laquelle est lue par 4 informateurs tous originaires du Sud Vietnam.

Les mesures sont faites sur le martinoscope du Laboratoire de Phonétique de l'Université Paris VII et des mingogrammes sont tirés par le détecteur de mélodie du Laboratoire de Phonétique de Paris III.

Sont retenues pour l'étude les données acoustiques (durée en cs, intensité en dB, Fo en Hz) relatives à la finale des syllabes-tests, à savoir le noyau vocalique et éventuellement la consonne nasale ou semi-consonne finales.

II. EXISTENCE D'UN ACCENT.

L'analyse des données acoustiques des syllabes d'après leur position dans les mots composés a permis de mettre en évidence 3 constantes:

1. La durée.

Indépendamment de la structure du noyau syllabique et du ton porté par la syllabe, sa durée en position de finale de mots est toujours plus longue qu'en postion non finale. Le rapport D(f)/D(nf) varie entre 1,53 et 2,08 autour d'une moyenne de 1,71.

Tableau 1 : Durée moyenne des finales des syllabes selon le type de mot composé (en cs).

Type de mot	Position				
	1	2	3	4	5
Dissyllabe				15,43	25,95
Trisyllabe			13,73	15,27	23,47
Tétrasyllabe		13,17	14,35	14,20	25,71
Pentasyllabe	14,58	15,25	13,66	11,50	24,00

2. L'intensité.

- Pour les mots comprenant 2 ou 3 syllabes, la tendance générale est à un léger accroissement du niveau d'intensité de la syllabe finale.
- Pour les mots de plus de 3 syllabes, l'intensité augmente en début de mot pour décroître ensuite jusqu'à l'avant-dernière syllabe puis remonter en fin de mot.

Tableau 2 : Intensité maximale des syllabes selon leur place dans le mot (en dB).

Type de mot	Position dans le mot				
	1	2	3	4	5
Dissyllabe				38,29	43,00
Trisyllabe			41,15	39,57	43,00
Tétrasyllabe		43,45	42,91	41,50	43,00
Pentasyllabe	43,37	44,62	42,62	41,25	43,00

3. La fréquence fondamentale (Fo).

La comparaison des valeurs du Fo d'une même syllabe qui se trouve dans différentes positions d'un mot composé, met en évidence le fait qu'en position de finale de mot, la courbe du Fo de la syllabe étudiée évolue dans un espace beaucoup plus étendu que lorsque celle-ci se trouve à l'intérieur ou au début d'un mot.

Autrement dit, dans le premier cas, il y a plénitude de la rélisation tonale, et dans le second, réduction de cette dernière.

Evolution du Fo (en Hz) de la syllabe MÁY dans 3 mots composés.

:Mot-test	Informateur 1		Informateur 2		Informateur 3		Informateur 4	
	Debut	Fin	D	F	D	F	D	F
xe máy	128	179	159	217	215	300	285	375
máy do	113	142	142	196	201	240	254	335
máy ti vi	110	146	137	190	201	263	233	332

4. Les marques de l'accent en vietnamien.

Selon Hoàng Tuê & Hoàng Minh (voir Etudes Vietnamiennes n°40 - 1975), il existe en vietnamien un accent dont les marques "positives" seraient un surcroît de durée et d'intensité, accompagné d'un maximum de plénitude dans la réalisation tonale. Cette étude confirme ces hypothèses dans l'ensemble. Toutefois, à mon avis, la plénitude tonale est une conséquence naturelle de l'accroissement de durée et d'intensité et par conséquent, ne peut être placée sur le même plan d'égalité.

Ainsi, dans un mot composé en vietnamien, il existe un accent qui tombe sur la dernière syllabe du mot. Cet accent est purement physique et dans bien des cas, les locuteurs natifs n'en ont pas conscience.

III. INFLUENCE DE L'ACCENT SUR LES REALISATIONS TONALES.

Dans le parler du Sud du Viet Nam, les tons sont au nombre de 5, répartis en 2 registres: 2 tons hauts NGANG (égal) et SẮC (montant); 2 tons bas HUYỀN (descendant) et NẶNG (descendant-montant); et un cinquième ton qui se déploient sur les 2 registres.(voir Doc.1)

En position accentuée (c'est-à-dire en finale de mot), l'évolution du Fo se fait de façon normale, et la courbe représentative de chaque type de ton est identique à celle des mêmes tons en syllabe isolée (voir Gsell-1980 et Seitz-1987). La présence de l'accent contribue d'autre part àu maintien d'une certaine stabilité au ton qui résiste mieux à l'assimilation des tons environnants.

En l'absence de l'accent, on peut donc remarquer des faits suivants:

1. Réduction de l'espace d'évolution des tons.

Les tons inaccentués se déploient toujours dans un espace tonal plus réduit, cela indépendamment de leur registre ou de leur contour mélodique. Cette réduction, d'après les données du corpus, varie de 30 à 50% selon les tons.(voir Doc.2)

Cependant, malgré cette réduction, le registre propre à chaque ton reste maintenu, c'est-à-dire que l'espace dans lequel évolue un ton peut se rétrécir, en l'absence de l'accent, mais il reste toujours distinct de celui d'un autre ton.

2. Instabilité du dessin mélodique.

Les tons en position inaccentuée sont plus sujets à l'assimilation du ton qui le précède, comme le montre le tableau suivant des intervalles entre les valeurs du Fo des tons après sac et après huyền:Les intervalles sont presque toujours plus grands quand le ton est en position inaccentuée que dans le cas opposé.(Doc.3)

Cette instabilité se traduit aussi par une certaine modification du contour des tons, surtout pour ceux à dessin mélodique complexe comme hoi et nang qui, sous l'inluence d'un ton bas qui les précède, peuvent perdre sa partie descendante pour devenir simplement montants.(Doc.4)

IV. COARTICULATION TONALE EN VIETNAMIEN.

Ce problème a été traité par Han&Kim (1974) et par Gordina & Bystrov (1984) mais pour les tons du parler du Nord.

1. La coarticulation tonale en vietnamien est fondamentalement progressive.

Un ton subit plus l'influence du ton qui le précède que celle de celui qui le suit. L'analyse du Fo des tons dans différents contextes montre qu'au début du ton les écarts sont toujours plus grands qu'à la fin.

D'autre part, après un ton haut, un ton débute généralement à un niveau plus élevé que d'habitude, et après un ton bas, à un niveau inférieure. Si l'on considère qu'après NGANG, un ton prend son départ à un niveau normal, voici en quarts de ton, les écarts de niveau par rapport à cette base dans d'autres environnements:

après SẮC: + 2 après HỎI: + 1
après NẶNG: -1 après HUYỀN: -2

2. La coarticulation tonale ne se propage pas de façon uniforme.

En général, les tons subissent des modifications au début de leur courbe et non à la fin. La comparaison des variantes des tons dans divers contextes montre qu'elles se différent par leur début et non par leur finale .

Tout se passe comme s'il existe une forte tendance à la conservation des caractéristiques distinctives du ton qui neutralise les différences contextuelles au cours de la réalisation tonale.

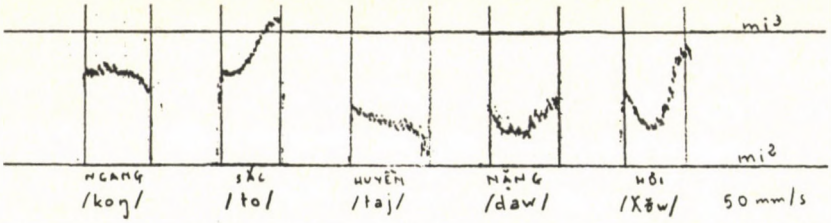
Cela explique pourquoi la coarticulation en vietnamien est progressive et qu'un ton est toujours reconnu comme tel par les locuteurs natifs.

D'une certaine manière, cette étude de la coarticulation tonale confirme la thèse de HOANG CAO CUONG (1985) selon laquelle un ton en vietnamien comprend toujours deux parties: une partie initiale, sujette aux influences contextuelles et qui subit diverses pressions de l'intonation, et une partie finale, plus stable, qui contient les informations sur les caractéristiques distinctives du ton.

REFERENCES

1. DOAN THIEN THUAT (1980) – Ngữ âm tiếng Việt – Dai Hoc & THCN – Hà Nội.
2. EARLE, M.A. (1975) – An acoustic phonetic study of Northern Vietnames tones – Speech Communications Research Laboratory Inc. – Santa Barbara – California.
3. FONAGY, I. (1983) – La vive voix – Payot – Paris.
4. GARDE, P. (1968) – L'accent – PUF – Paris.
5. GORDINA, M.V. & BYSTROV, I.S. (1984) – Foneticheskii stroi vietnamskogo jazyke – Izdatelstvo "Nauka" – Moscou
6. GSELL, R. (1979) – Sur la prosodie du Thai standard: tons et accent – Université Paris III – Paris.
   (1980) – "Remarques sur la structure de l'espace tonal en vietnamien (parler de Saigon)" dans Cahiers d'Etudes Vietnamiennes – Université Paris VII – Paris.
7. HAN, M.S. & KIM, H.O. (1974) – "Phonetic variation of Vietnamese tone" dans Journal of Phonetic 2 –
8. HOANG CAO CUONG (1985) – "Bước đầu nhận xét về đặc điểm ngữ điệu tiếng Việt" dans Ngôn Ngữ 3 – Hà Nội.
9. HOANG TUE & HOANG MINH (1975) – "Remarques sur la structure phonologique du vietnamien" dans Essais Linguistiques – Etudes Vietnamiennes 40 – Hà Nội.
10. ROSSI, M. (1981) – L'intonation: de l'acoustique à la sémantique – Klincksieck – Paris.
11. SEITZ, P. (1986) – Relationships between tones and segments in vietnamese – University of Pennsylvania – UMI – Ann Arbor – Michigan.
12. TRAN THIEN HUONG (1976) – Interrogation et Intonation en vietnamien – Mémoire de maîtrise – Université Paris VII – Paris.

Document 1: Tons en vietnamien (parler du Sud)



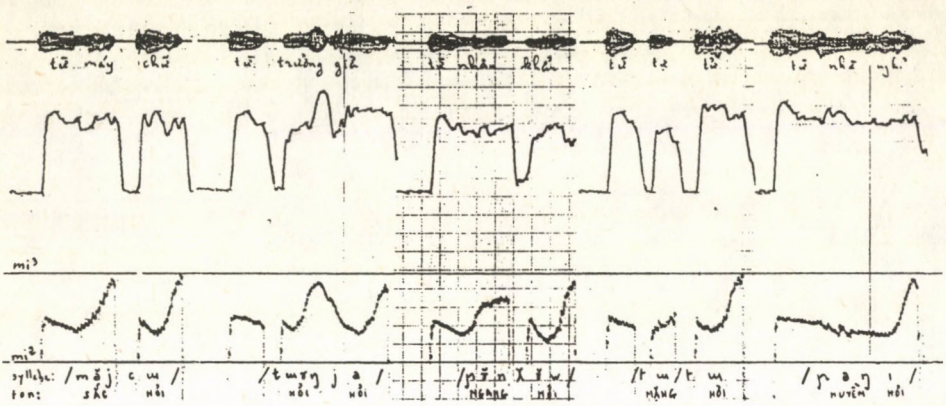
Document 2: Fo des tons en position accentuée et inaccentuée.

Informateur	Ton	Accentué: Fo moyen	Evolution de Fo			ΔF	Fmax./Fmin.	Inaccentué: Fo moyen	Evolution de Fo			ΔF	Fmax./Fmin.
I	SAC	141.5	128		174	46	1.35	140.5	127		158	31	1.24
	HUYEN	109.5	112.5		98	-14.5	1.14	107	110		101	-9	1.08
	BANG	128	130.5		124.5	-6	1.04	128	131		125.5	-5.5	1.04
	HOI	113	105	95.5	149	53.5	1.56	115.5	112.5	107	134.5	27.5	1.25
	NANG	105	112.5	97.5	115.5	18	1.18	109	112	102.5	111	8.5	1.08
II	SAC	167	155.5		213.5	58	1.37	175	164		202	38	1.23
	HUYEN	133	145		124	-21	1.16	141	148.5		135	-13.5	1.10
	BANG	159	167.5		156.5	-11	1.07	164	167.5		162	- 5.5	1.03
	HOI	141	137.5	126	206	80	1.63	142	141.5	131	164	33	1.25
	NANG	130	143	124.5	151.5	27	1.21	134.5	138.5	128	144.5	16.5	1.12
III	SAC	226.5	207.5		296	88.5	1.42	236	215		275	60	1.27
	HUYEN	169.5	189		149	-40	1.26	175.5	182		167.5	-14.5	1.08
	BANG	216 .	217		200.5	-16.5	1.08	216.5	219.5		214	-5.5	1.02
	HOI	182	170.5	153.5	261	107.5	1.70	185.5	184	164	218	54	1.32
	NANG	162.5	176.5	155	176	21	1.13	166	170	155	173	18	1.11
IV	SAC	301	276		365	89	1.32	303	278		337.5	59.5	1.21
	HUYEN	221.5	246		201.5	-44.5	1.22	228.5	244.5		217.5	-27	1.12
	BANG	277	287.5		271	-16.5	1.06	286.5	287		279.5	- 7.5	1.02
	HOI	237.5	237	209	317.5	108.5	1.51	242	244.5	217.5	288.5	71	1.32
	NANG	221	232	207.5	248	40.5	1.19	231.5	234.5	219	248.5	29.5	1.13

Document 3: Ecarts de fréquence entre les débuts des tons après sac et huyên.

TON	Informateur1 Inacc.	Acc.	Informateur2 Inacc.	Acc.	Informateur3 Inacc.	Acc.	Informateur4: Inacc.	Acc.:
SAC	1.24	1.15	1.26	1.13	1.18	1.11	1.18	1.17
BANG	1.23	1.13	1.13	1.12	1.17	1.13	1.08	1.16
HOI	1.08	1.14	1.17	1.05	1.10	1.04	1.15	1.05
HUYEN	1.15	1.14	1.13	1.16	1.02	1.01	1.11	1.08
NANG	1.30	1.16	1.27	1.20	1.13	1.02	1.13	1.10
MOY.	1.20	1.14	1.19	1.13	1.12	1.06	1.13	1.11

Document 4: Modifications du ton HOI (accentué) après les 5 tons du Sud .



Document 5: TABLEAU RECAPITULATIF DES VALEURS DES TONS DANS DIFFERENTS CONTEXTES (Informateur 4).

ACCENTUEE

Ton	Après	Moyenne des différentes courbes Moy.	Début		Fin	Nombre d' occurences
SAC	sac	318	298		370	2
	hoi	306.3	284.6		366.6	3
	bang	296.8	279.4		357.6	5
	nang	289.5	272		339.5	2
	huyên	284	254		359	1
BANG	sac	284	308.5		279	2
	hoi	284.5	293.5		275	2
	bang	276.7	280	289.8	259	7
	nang	273	275.3	289.3	262.6	3
	huyên	269.5	269	285.5	263	2
HOI	sac	237	225	205.5	312.5	2
	hoi	237	233	210	313.5	2
	bang	231	243.5	202	316	2
	nang	236.5	225.7	214.5	323	4
	huyen	238	230.5	212	321.5	2
HUYEN	sac	227	255		201.5	2
	hoi	217.5	243		204	2
	bang	214	233		201	1
	nang	218	246		201	1
	huyên	211.5	236		201	2
NANG	sac	227	239.3	205.6	248.6	3
	hoi	229.5	245	209	252	2
	bang	217	232	199	243.5	2
	nang	209	225	198	230	1
	huyên	201	207	196	225	1

INACCENTUEE

Ton	Après	Moyenne des différentes courbes Moy.	Début		Fin	Nombre d' occurences
SAC	sac	320	300.5		342	2
	hoi	317.5	294.5		333	2
	bang	310	293.5		343	2
	huyên	284.5	254.4		332.4	19
BANG	sac	294.2	300.8		287.4	5
	bang	281.7	280.4	285.7	273	7
	huyên	278.8	270.9	286.8	272.2	18
HOI	sac	247	256.5	219	286	2
	bang	235.7	242.5	220.2	281	4
	nang	233	225	215	280	1
	huyên	236.1	220.7	212	296	9
HUYEN	sac	234	258		218	2
	huyên	223.2	231.3		217.4	10
NANG	sac	237	248.6	218.3	250.6	3
	hoi	234	240.6	212.6	244.6	3
	huyên	227	223.6	219.6	244.6	6

ON THE PHONETIC ANALYSIS OF THE SPOKEN TEXTS

Éva FÖLDI
Department of Phonetics, Loránd Eötvös
University, Budapest, Hungary

Introduction

Research in textology has recently gained renewed impetus in Hungary. The investigations are focussed on defining the notion of text, as well as various aspects of its syntactic, grammatical, and phonetic analysis.

In my paper I present the phonetic analysis of the suprasegmental structure of spoken texts, based on a recording of a literary piece.

It is well know that speech can be divided into segmental and suprasegmental structures. The former is created by joining together the speech sounds of a given language according to some definite rules. This is way larger linguistic units of language (syllable, word, sentence etc.) are formed. Thus the minimal unit of the segmental structure is the speech sound. Suprasegmental structure means the prosodic composedness of speech, the smallest units of which might be called prosodemes. The text -- a larger sense unit -- can minimally be segmented into sentences syntactically and into phonetic phrases phonetically. A phonetic phrase is a suprasegmentally composed, distinct unit of speech text (5).

Suprasegmental structure is used here as a synonym for what is traditionally called intonation or speech intonation. I use the former term to suggest the complexity of this phonetic category, i.e. the fact that so-called speech intonation is a linguistically relevant phonetic subsystem consisting of a number of acoustic parameters such as

1. melody (tone) -- an element based on the changes of fundamental frequency,
2. dynamics
3. pause } the components which are based on the changes of intensity
4. tempo
5. rhythm } parameters, based on the changes of duration
6. emphatic lengthening
7. timbre -- formed by the number and the relative force of formants. (2, 4)

In the Hungarian language suprasegmental structure is relevant on the level of sentences. Its primary functions are delimitation and modality.

Methods

For the instrumental analysis I have chosen a part of Sándor Csoóri's prosaic work entitled "Nagy László földi vonulása" ('László Nagy's Earthly Proceeding') and now I repre-

sent the acoustic projection of the suprasegmental structure (in a broader sense -- intonation) of this literary text received by way of experimental phonetic analysis. I used Kálmás Bolla's methods in the analysis (1). The text was recorded in the pronunciation of a male informant (P1), whose register, i.e. the fundamental frequency band in which the text was realized was 75--175 Hz. The acoustic diagrams and registrations were made by FFM 650 fundamental frequency meter, IM 360 intensity meter and 34T four-channel mingograph. I made the synthesis by a software for Commodore 64 (3). after giving the data of the diagrams I also made the phonetic transcription of the suprasegmental structure of the text.

Results

From the syntactic point of view the text consists of 20 simple, compound, complex and composite affirmative sentenses and two citated lines, phonetically it contains 52 phrases. They are divided within the whole text as follows:

1. name of the author and title of the work -- 3 phonetic phrases,
2. introductory part (1 paragraph) -- 5 sentenses and 10 phonetic phrases,
3. first paragraph of exposition -- 4 sentences, 9 phonetic phrases, second paragraph (the essence of the so-called "message") -- 10 sentences and 20 phonetic phrases, the cited lines -- 5 phonetic phrases,
4. ending part (1 paragraph) -- 1 sentence and 5 phonetic phrases.

The total time of the text is 168215 msecs, from which the pauses take 43729 msecs. The duration of the phonetic phrases varied from 540 to 8600 msecs. The pauses at the sentence boundaries and between the phonetic phrases took 405--1820 msecs and 120--985 msecs, respectively. Thus the pauses segmenting the text and those expressing emotions or emphasis were realized in a quite wide range of duration of 120--1820 msecs. (The breath pauses are naturally also included.)

I represent the tempo of speech by the number of sounds per seconds. I measured 11.3 sounds/s as the average tempo, which corresponds to the tempo of an unbiased lyrical text consisting of affirmative sentences.

The rhythm of the text formed by the relative duration of syllables -- due the above-mentioned stylistic factors -- was relatively steadily pulsating while in longer phonetic phrases and those containing new information it was slightly 'cracking'.

The lowering tone that is the typical means of intonation for expressing affirmation and completedness in Hungarian speech was characteristic of the tune of the whole text. The falling tone was most frequently realized in the following variations: rising--falling, rising--falling--rising--falling (tremolo), rising--floating--falling, falling--rising--falling, rising--falling--floating--falling. The phonet-

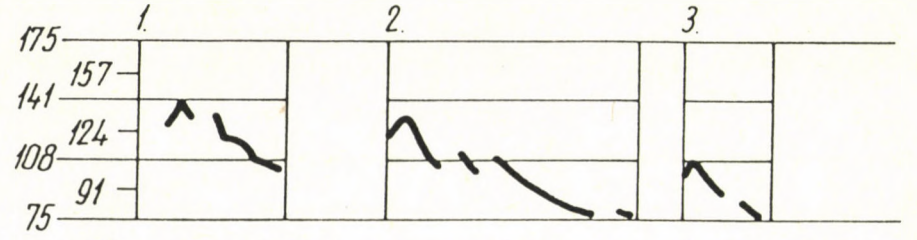
ic phrases containing enumeration or preceding a continuation usually had a floating or slightly rising tone. From the tessitura, register and interval data of the 52 phonetic phrases it can be seen that the relatively low fundamental frequency peak values, characteristic of affirmative sentences alternated between 90--175 Hz, while the Fo's took a minimal value of 75--100 Hz. The average interval of P1 was 38%. The phonetic phrases were usually realized in a low--mid tessitura band.

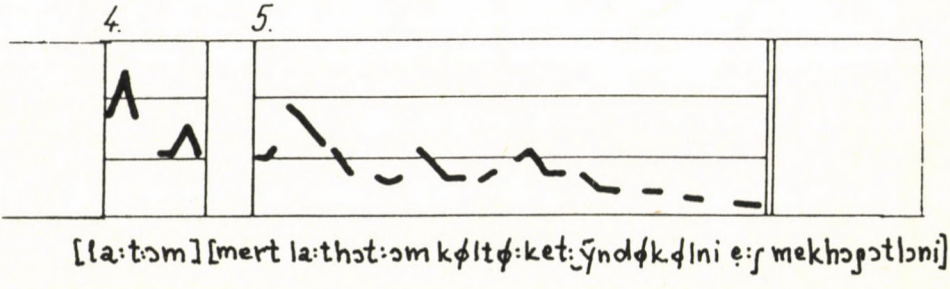
The intensity values forming the dynamics of speech were between 10--40 dB. The peak values were 30--40 dB, and they frequently occurred along with the fundamental frequency peak values. The minimal values of intensity were 10--28 dB, and in most cases were measured at the end of phonetic phrases expressing completedness. The intensity and Fo peak values had an emphatic role and they usually occurred at the beginning of the phonetic phrases.

According to its contents and message the text did not contain emphatic lengthening.

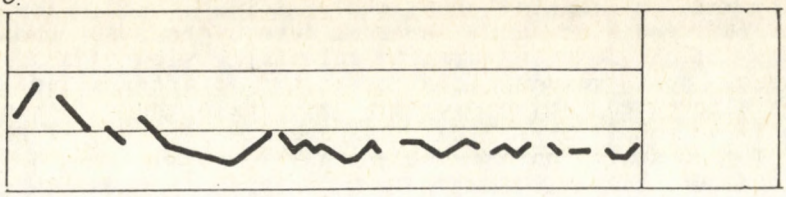
I completed, checked and tested the analysis by synthesis, employing a software for Commodore 64 (3). This program only sounds the suprasegmental structure, and monitors the intonogram of the structure as well. In this figure one can see the Fo changes of the tune as a function of time. One can also see the tune diagram represented in low, mid and high register bands, the total time and the values of tempo, volume and interval calculated by the computer. It is possible to repeatedly voice the suprasegmental structure, so the visual experience is reinforced by an auditive one as well.

Here I am going to give the Fo pattern of a few phonetic phrases.



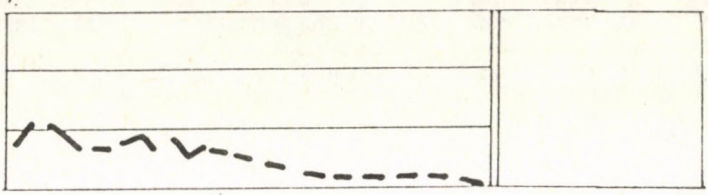


6.



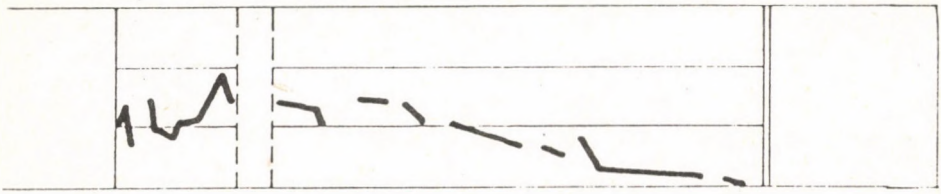
[la:t:ɔm ɸ:ket vila:glɔpok elʃɸ: oldɔla:n filmen tɔla:lkozo:kon]

7.



[ʃ la:t:ɔm ɸ:ket rõntʃtelepek orfeusɔkē:nt elʃodro:dni]

14



[ę:ʃ kɸzbeŋla:t:ɔm] [mert la:thɔt:ɔm nɔʒ la:slo: fɸldi vonula:ʃa:t]

References

1. BOLLA Kálmán : Apáczai Csere János : Az iskolák felettébb szükséges voltáról (Részlet). Elemzés. EFF 1. 1988, 55--69.
2. BOLLA Kálmán : A beszédfolyamat intonációs elemzése és az intonáció fonetikus lejegyzése. MFF 3. 1979, 19--31.
3. BOLLA Kálmán--PAPP György--RADVÁNYI Péter : Beszédintonáció. Számítógépes program C=64-re a beszédintonáció tanulmányozására. ELTE Fonetikai Tanszék, 1987, 14 p.
4. FÖLDI Éva : A kérdés kifejezésének intonációs eszközei a magyarban és a lengyelben. MFF 5. 1980, 109--17.
5. ROPA, A.--RUSOWICZ, A. : Rola cech prozodycznych w segmentacji tekstu mówionego. In : Z zagadnień fonetyki i fonologii współczesnego języka polskiego. (Księga referatów ogólnopolskiej konferencji w Toruniu 1978 r.) Toruń, 1982, 119--26.

TONE-SPECIFIC PATTERNS OF LARYNGEAL CONTROL IN CHINESE

Pierre HALLE*, Seiji NIIMI**, Satoshi IMAIZUMI**
*Laboratoire de Psychologie Expérimentale,
CNRS-EHESS, Université PARIS V, Paris, France.
**Research Institute of Logopedics and Phoniatrics,
University of Tokyo, Japan.

INTRODUCTION

In Modern Standard Chinese, tones are produced mainly by means of laryngeal articulatory gestures, rather than by the control of subglottic air pressure. A preliminary Electromyographic investigation of Cricothyroid and Sternohyoid muscle activity during the phonation, suggested that laryngeal gestures involved in the production of tones were controlled by rather stable and tone-specific CT activity patterns. Such clear patterns were observed for SH, only in the case of the low-dipping tone T3. Other possible tone-specific SH patterns were blurred out by the strong involvment of SH in supralaryngeal gestures.

We report here the results of a second EMG experiment, where the speech material had segmentals that minimized SH muscle involvement in supralaryngeal articulation (minimal jaw opening and/or tongue backing). CT and Vocalis muscle activities were also recorded. The main finding, which upholds the hypothesis of tone-specific activity patterns, is that SH muscle activity systematically occurs at the onset of tone T2, as well as at the offset of tone T4. In contrast, SH muscle activity is very weak in the case of tone T1. The observed activity for tones T2 and T4 should thus be related to the laryngeal articulation. Secondary results deal with Vocalis and SH activity at the end of breath groups.

To summarize, we are now able to give a more detailed account of the canonical motor activity patterns underlying tone production in Modern Standard Chinese.

PREVIOUS RESULTS

In previous experiments conducted in Paris (with Pr. Chevrie-Muller, INSERM, see [1]), we gathered CT activity data for one Chinese female subject (from TaiPeh), and then, both CT and SH activity for another female subject (from Beijing). The speech material consisted of CV syllables embedded in a carrier sentence, and belonging to minimal series sharing the same segmentals at the 4 tones: /ma/, /wu/, /ai/, /du/, /tu/, /bi/, /ge/, and /fa/.

It appeared that CT activity was consistently correlated with Fo raising for both subjects, with a lead estimated to 80 ms. A sustained activity was found for tone T1 (high level tone), roughly fitted with the syllable rhyme duration, a more intense but shorter activity for both tones T2 (mid rising) and T4 (high falling), located at the end of the rhyme (T2), or slightly before its onset (T4), and finally, no consistent activity for tone T3 (low dipping). On the other hand, SH activity was found to contribute a

great deal in supralaryngeal gestures (jaw closing/opening, tongue lowering and backing), a rather classical outcome (J.Ohala & H.Hirose, 1970). According to our data, the role of SH in Fo lowering - or at least, in helping Fo lowering - was systematic in the case of tone T3, but not T4.

In the view of these preliminary results, we felt that the design of a more systematic corpus of speech material was needed, in order to minimize SH supralaryngeal involvment.

THE NEW EMG EXPERIMENT

Like in the previous experiment, we used Modern Standard Chinese syllables embedded in a frame sentence. The frame sentence was /yi ge X zi/ ([ʲik Xts], "A character X"), where X was a syllable belonging to mimimal series sharing the same segmentals at the 4 tones. Four minimal series were used: /yi/ ([ʲi]), /bi/ ([pi]), /mi/ ([mi]), and /hu/ ([xu]), designed in order to minimize supralaryngeal activity of strap muscles. Some supralaryngeal activity however, should appear in the case of [pi] (jaw closing/opening) and /hu/ (tongue backing). The corpus consisted of 4 x 4 = 16 sentences, each of them repeated 10 times. The acoustic onset of /ge/ was used for time-aligning utterances in the averaging process.

The subject was a male native speaker of Modern Standard Chinese, born and raised in Beijing, aged 26, with no known abnormality, wether in speech production or in speech perception.

The EMG activity of 3 muscles, CT, Vocalis, and one strap muscle, assumed to be SH, together with the audio signal, were recorded by means of a U-matic video recorder. Non speech activities allowed for controlling EMG electrodes positionning.

Fo was extracted from the digitized audio signal (sampled at 10 kHz) by means of a cepstral method.

Audio and EMG signals were rectified and integrated, yielding digitized signals with a sampling period of 2 ms.

The "line-up" event used for the averaging process was the acoustic onset of /ge/. The averaging was performed both for rectified/integrated signals and for Fo. Finally, the averaged EMG data were smoothed by a 10 points long (i.e. 20 ms) windowing. The following averages have been computed: per tone and segmental across repetitions (4x4 averages), and per tone across segmentals and repetitions (4 averages).

RESULTS

CT activity patterns confirm our previous findings: a sustained activity for tone T1, roughly fitted with syllable rhyme duration, a more intense and shorter activity for T2 or T4 at the end of the rhyme or slightly before its onset, and no significant activity for T3. In the case of T2 and T4 the time lag between the peak activity of CT and the peak Fo value can be estimated to about 90 ms. In the case of T1 activity, there is a well defined CT peak activity, but no well defined peak Fo value (T1 tone contour is basically even). However, the section of the contour where it

becomes flat after an initial rise is rather narrow (about 30 ms), and the time lag between CT peak activity and this region is also about 90 ms.The subsequent lower but steady CT activity should serve to maintain the Fo level reached at the rhyme onset.

Contrarily to our previous experiment, tone specific SH activity patterns consistently appear for each of the 4 tones. For T3, SH activity is always very intense and centered on the syllable rhyme. Unlike CT activity, the time lag is very small. The same contrast in timing behaviour between CT and SH has been observed by other investigators (H.Hirose & M.Sawashima, 1981). In the case of T2, there is a systematic SH activity centered on the rhyme onset. As for T4, there is also a systematic SH activity occuring slightly after the center of the rhyme. In the case of T1, SH activity is very weak and depends on the segmentals rather than on the tone. It is the weakest for /yi/ and /mi/ syllables, small but consistent before the release of [p] in /bi/ or during the friction of [x] in /hu/ syllables, but always much weaker than SH activity found in the case of T2 or T4 whatever the segmentals be.

The activity of Vocalis is always weak, though weakly correlated with CT activity. It may be the case that the speaker was keeping on with the same "register". However, some noticeable increase of Vocalis activity consistently appears at the end of each utterance (whose last syllable /zi/ is at tone T4), coinciding with a very intense SH activity.

DISCUSSION

The results suggest that stable patterns of - at least - CT and SH activity may be ascribed to each of the 4 tones in Modern Standard Chinese. These patterns are illustrated in Fig. 1 and are summarized as follows.

- tone T1: sustained CT activity, SH inhibited.
- tone T2: SH activity at the rhyme onset, then intense and short CT activity at the end of the rhyme.
- tone T3: intense SH activity centered on the rhyme, CT inhibited.
- tone T4: intense and short CT activity before rhyme onset, then SH activity slightly after the center of the rhyme.

What has been found in this study, is the existence of tone related SH activity for T2 and T4, thanks to the choice of syllables where SH supralaryngeal involvment was minimized. This appears clearly especially by reference with SH behaviour in the case of T1. The effect of the "tone" factor overrides the effect of the "segment" factor when tone is varied from T1 to T4, and segmentals are varied from /yi/ to /hu/. This is best illustrated by the different SH activity patterns observed for /yi/ at each of the 4 tones. Since CT activity is dominant in the case of T2 and T4, raising Fo during the rhyme (T2) or before the rhyme (T4), we call T2 and T4 SH activities secondary activities.

This finding is in good agreement with acoustic data from Kratochvil (1985). He found that the longer is a T2 syllable, the lower is its Fo contour onset, and a converse outcome for T4. In Modern Standard Chinese, duration is well correlated with stress, and we may assume that the more a syllable is stressed, the more intense should be the involved muscular activities. Thus a longer

T2 syllable should be produced with a more intense initial SH activity, resulting in a lower Fo at the rhyme onset. A similar interpretation holds for T4.

Finally, the strongly marked SH activity at the end of all utterances is probably related to the final sharp intonation downdrift as well as to the tone T4 of /zi/. Interestingly, this SH "downdrift activity" is accompanied by a markedly increased Vocalis activity. One interpretation is that Vocalis activity is related to the sharp downdrift somehow specific to this speaker who is usually terminating each breath group in a kind of vocal fry.

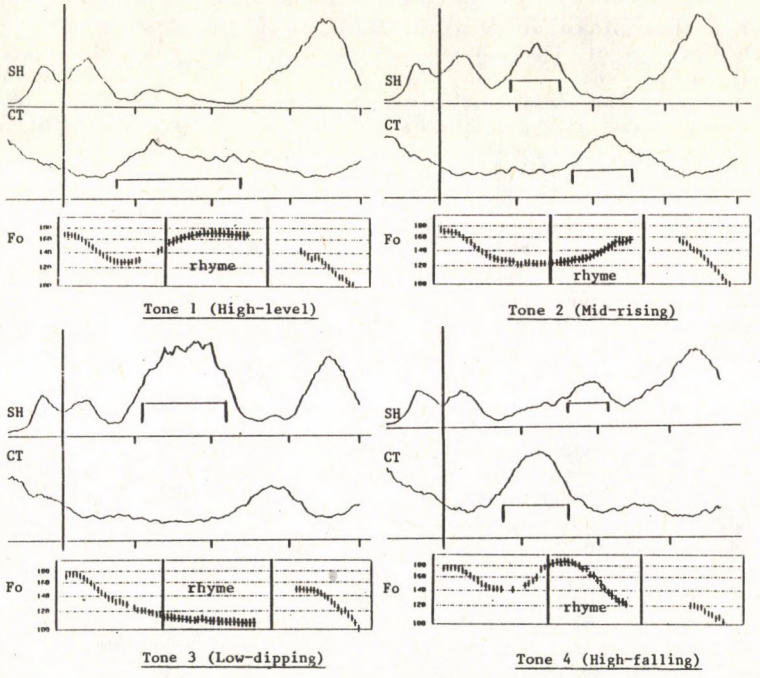


Figure 1: Tone-specific CT and SH activity patterns. The highlighted part of Fo contour marks the target syllable rhyme. The identified patterns are shown by the bracketed portions of EMG envelopes.

REFERENCES

[1] Sagart L., Hallé P., Boysson_Bardies B. & Arabia-Guidet C. (1986). Tone Production in Modern Standard Chinese: an Electromyographic Investigation. Cahiers de Linguistique Asie Orientale, XV (2).

[2] Ohala J. & Hirose H. (1970). The Function of the Sternohyoid Muscle in Speech. Ann.Bull.RILP, 4. pp. 41-44.

[3] Hirose H. & Sawashima M. (1981). Functions of the Laryngeal Muscles un Speech. In K.N. Stevens & M. Hirano (Eds.) Vocal Fold Physiology. Tokyo University Press, pp. 137-154.

[4] Kratochvil P. (1985). Variable Norms of Tones in Beijing Prosody. Cahiers de Linguistique Asie Orientale, XIV (2).

Contour and Accent Structure Shifts:
An Autosegmental Approach to Intonation Variants.

Anthony Hind
UFRL Université Paris VII

INTRODUCTION

I am presenting here a version of the autosegmental theory which differs from traditional autosegmental analysis both in incorporating the distortion theory of Fonagy [2] for interpreting attitudinal variation and in attempting to derive all dependant prenuclear contours from basic or non derived nuclear contours(Hind [6]).

I will first show for the intense falling and rise-falling contours how the distortion theory contributes to our understanding of attitudinal variants as significant distortions from expected target realizations (or phonotypes), predicted by the obligatory rules from underlying segmental contours. Then, I will discuss the role of nuclear peak delayal in attitudinal and dialectal variation:

I INTENSE VARIANTS

Phonetic and semantic data which came to light during psycho-phonetic tests, brought me to distinguish two contrasting phonological contours underlying Falling phonetic contours. These as in (1) are roughly equivalent to the traditional Falling and Rise-falling contours.

(1) H L +L (E), and L H +L (R-E),

This clearly differs from the approach in Gussenhoven [4] and Ladd [8] which treats the rise-falling contour as an attitudinal variant of the falling intonation obtained by nuclear peak delayal from a single underlying *High + Low* falling phonemic contour. For example, in (2a) below, the neutral Falling contour is obtained first by the association of the contour to the text (with the "high nuclear tone" associated, by *copy*, to all the pre-nuclear accented syllables); this is, then, followed by the application of Down-drift, which lowers each successive high tone by one degree, giving a Falling stepping head:

(2a) They noticed him waiting in a cinema queue #
H H^{-1} H^{-2} L L

Now in my theory, Down-drift can be formulated so as to be blocked before a certain type of pause, or rupture, (noted here with a square pause sign); when this is placed, optionally, before the nuclear syllable as in (2.b):

(2b) They noticed him waiting in a □ cinema queue #
H H^{-1} H^{-0} L L

The association of this pause before the word containing a nuclear syllable results in an abnormally high nuclear syllable which is very probably judged as a "distortion" in relation to the phonotype of (2.a), (i.e. +2 degrees) which could either express a tense attitude on the part of the speaker(amazement, surprise, etc. via increased vocal tension) or simply be a means for focussing the attention of the addressee on the word containing the nuclear syllable (special stress in the traditional theory).

II Accent and Pitch structure misalignment:

A) High tone displacement and exageratedly surprised(or ironic) contours:

In R.P., the Rise-falling contour, on the other hand, has a rising stepping head: the low tones associated to the pre-nuclear stressed syllables by copy from the low initial tone of the contour are affected by updrift; this process can also be blocked before the nuclear syllable for the purpose of expressiveness or contrast. In this case, however, it gives an abnormally low departure of the nuclear syllable(-2 degrees) as in (3).

However, (3) has in fact undergone a further distortion through the application of an optional High tone displacement rule which moves the nuclear peak tone to the left in HL contours, and to the right(as here) in LH contours.

```
                    +          +        +        *
3)  He  thinks  they're going to elect a □ republ ic an #
                                                 *
                    L          L+1      L+2      L+0   H    L
```

In such cases, increased speaker intrusion is manifested by the disturbed accent/pitch structure giving an exaggerated intense effect suitable, according to context, for the expression of extreme surprise or irony.

B)Rythmic(predictable: Fónagy[3]) and highly stylized contours:

However, some cases of tone prominence on non-accented syllables even in R.P. do not give rise to "surprised/emphatic contours". In certain rhythmic realizations - where the metrical grid imposes strict alternating strong/weak patterns - a weak, or even reduced, vowel can be aligned with a strong metrical position, and have a high or low tone associated to it, as is the case with the final vowel of Canada in (4):

```
          +          *
4)  Cana§ da is   § gre en #
     x    x         x
     x x  x  x      x x
     L H  L  H      H   L
```

(where "§" is a rhythmic juncture).

This realization while being emphatic, is nevertheless highly predictable in the position of the tones, and is most probably found as a desperate repetition "just for the form" (I've told you before, but I can see I will have to tell you again). It can never be used to express surprise from the very nature of its predictability.

The tendancy to trochaic structures in particularly melodic or rhythmic productions by professional speakers and actors has often been noted in English. A recorded extract from 'The Care taker' clearly shows this. Counter to the English stress rules, an actor forces a trochaic rhythm on to each word of a streatch of text, repeatedly aligning the high nuclear tone with the initial syllable of words even when they contain a reduced vowel(approximately as in (5)); however, predictability, here, is more part of the music of the rendering(that is to say it relates to the conventional nature of theatrical performance) than due to the stereo-typed nature of the exchange. The degree of distortion from stress governed realizations which this produces, seems to be used by the actor to express strong (if stylized) emotion.

```
        +        +        +        +        +        +        *
5) Experienced first-class professional interior and exterior decorater
   x        x        x        x        x        x        x
   x x x x  x  x     x x x x  x x xx x  x  x xx  x x x x
   H L      H  L     H L      H L      H  L     H   L
```

To summarise, I have claimed that there would be a close relation between predictability(or regularity) of contour and the conventional(or stylised) nature of the communication. Whereas the more the contour differs from that predicted by the obligatory rules for associating intonation contours or from the predictions of strict application of the metrical grid, the greater the speaker intrusion in his text.

III DIALECTAL AND INTERLANGUAGE VARIATION

The distortion theory also correctly predicts that speakers of dialects which differ in their contour accent/structure association, will tend to be received by each other as deviating significantly on the attitudinal level(ie their utterances will seem to continually vehicle secondary messages of the same type, when judged from the standpoint of the speaker of the other dialect), this may well at least partly explain certain stereotype characteristics which one linguistic group tends to attribute to another.

'owever, an argument against my analysis of falling and rise-falling contours made from the case where a contour perceived as falling in one dialect or 's replaced (with apparently the same function) by a contour perceived as

rise-falling in another dialect or language. According to the 'single falling contour hypothesis', if both contours were basically HL in their underlying forms then the rise-falling contour could be derived from the underlying falling contour by the effect of nuclear peak delayal(Gussenhoven [4]). One possible candidate for this type of treatment could be the difference in use of the Rise-falling contour in R.P. and in Welsh English intonation. In Welsh English rising level or rise-falling contours, are frequently used as the neutral contour, where RP would use a fall. In Hind [7] I have already spoken of Scottish English so let us examine the case of Welsh English variants. The words: *alliances*, *father*, and *valid*, of a Welsh regional speaker have rise-falling contour as in (6a), compared to the falling contour used by an RP speaker in (6b):

```
                *                *               *
6a)  alliances          father          val id
       LH  L             L  H  L         L  H  L

                *                *               *
6b)  alliances          father          val id
       HL                H  L            H  L
```

It is probable that these Welsh Rise-falling realizations will give the southern RP speaker the impression that the Welsh speaker is always over-emphatic, taken to flights of fancy and exageration. Interestingly, on the other hand, I recently heard a southern English speaker on BBC4: "Midweek" claim that his Falling RP contours had him classed as "dogmatic" by other dialect speakers he encountered. This is exactly what the distortion theory would predict.

Now, Gussenhoven could claim that the optional High tone realignment rule of RP with which he derives the rise-fall as an intense variant of the underlying falling contour is an obligatory rule in Welsh English giving a rise-falling(with post nuclear high tone as in (6a)), or rising level neutral contour according to how far the nuclear High tone has been displaced to the right. However, I think the facts described in Williams[9](concerning Welsh contours) coincide remarkably with the data described in Hind [7] concerning some French student's productions of English contours which can't be explained within the High tone delayal hypothesis. These facts, involving inter-language influence, can be briefly resumed as follows:
French students learning English systematically produce a rise-falling contour, when repeating a word such as *terrifying*, or *scientist*, as spoken by an RP speaker with falling intonation. The intonation peak in this case falls on the post-nuclear syllable as in (7a) contrasting with (7b) for the RP speaker.

```
7a)        *                 *
        terr i fy ing     sci en tist
         L   H  H  L        L H   L
7b)        *                 *
        terr i fy ing     sci en tist
         HL          L       L H   L
```

This suggests that nuclear High tone delayal is the valid explanation. However when you point out to a speaker(using a visual pitch analyser) that the English contour for the word 'terrifying'(7b) goes down on the first syllable, but stays down over the whole word, the French speaker very generally stays down over the first two syllables; rising, however, on the pre-final syllable as in (8):

```
8)                  *
                terr i fy ing
                 L   L H   L
```

It is difficult to consider nuclear peak delayal as part of a strategy for keeping the contour low over preceding syllables(for more arguments see Hind[7]).
Now, in certain emphatic forms of French intonation the initial syllable of a word is lowered and yet a normal Falling contour effects the end of the word. This complex contour seems to be made up of an LH associated to the beginning of the word(in dissimilation) with the final nuclear HL, associated to the end of the word, as with "terrifiant" in (9):

```
(9)             terr if  i ant
                 L   H   H  L
```

This is very close to the realization in (7a) for "terrifying". And I suggest that when the English contour is non final a French speaker tends to add the French final HL contour to the following text; but the rules of French in turn impose on him an LH contour for the preceding English nuclear syllable(Two following HL contours within the same phonological phrase would form an intonation clash according to the rules of French) and the preceding contour would undergo iambic reversal as in (10)

(10) Iambic reversal:

HL + HL → LH + HL

An explanation of this sort could both account for the similarity between Welsh and French English contours - compare the Welsh English realization of *alliances* in (11a), with the French realization of *scientist* in (11b) - and would be compatible with William's remarks about Welsh language contours.)

11a) * (11b) *

alliances (Welsh) sc i en tist (French)

LH (L) L H L

The realisation(11a) appears to have features in common with the Welsh language contours described in Williams[9]. According to Williams "In Welsh polysyllables, it is always the penultimate syllable that is stressed ..."; "...however , even if this syllable can be pitch prominent it ... usually steps down rather than up."... "The final unstressed syllable on the other hand ... is usually higher containing a pitch glide...". Unsurprisingly, for English informants these final syllables appear stressed.

Williams explains what she calls "this strange behaviour of stress in Welsh" by the historical development of an "Old Welsh Accent shift" which in the eleventh century, shifted word stress in Welsh from the final to the penultimate syllable; possibly, leaving behind a pitch accent on the final syllable. There would be rhythmic stress on the penult, and pitch prominence on the final syllable.

Thus the Welsh, and Welsh English contours, are compatible with the French English analysis: a final contour being associated with the last syllable according to one set of language principles and then another contour being associated to the most stressed syllable according to a different principle(possibly with Iambic reversal).

This seems to confirm the suggestion I made in Hind [5] and [7] that dialects of English with Rise-fall neutral contours such as Scottish and Welsh English could have developed from interlanguage interference(cf. also Cruttenden [1]:85.3.1.).

References:

[1]Cruttenden, A. «Intonation», C.U.P., (1986).

[2]Fónagy I. «La vive voix, essais de psycho-phonétique», Payot (1983).

[3]Fónagy I. et al. «Clichés mélodiques», Societas linguistica Europea, p. 273-303, (1983).

[4]Gussenhoven C. «On the grammar and semantics of sentence accents», Dordecht:Foris, (1984).

[5]Hind A. «Research on English intonation in an autosegmental framework», C.E.L.D.A.: le Suprasegmental, Université Paris Nord: Villetaneuse, avril (1984).

[6]Hind, A. «Phonosyntaxe: Place et Fonction de l'Intonation dans une Grammaire» . Thèse de Doctorat d'Etat, non-publiée, Université Paris VII, (1986).

[7]Hind, A. «Attitudinal and dialectal variation in intonation: High tone displacement and the role of the distortional component in Autosegmental theory» Proceedings of the 11th Congress of Phonetic Sciences, Tallinn ,USSR. (1987)

[8]Ladd, D.R. «On Intonational Universals», The Cognitive Representation of speech, T. Myers et al. eds., North Holland Publishing Company, (1981).

[9]Williams B. «An Acoustic study of Some Features of Welsh Prosody» In C. Johns-Lewis Ed. Intonation in Discourse Biddles Ltd, Guilford & King's Lynn GB (1985).

AUTOMATIC RECOGNITION OF PROSODIC CATEGORIES

David HOUSE
Department of Linguistics and Phonetics
Lund University, Lund, Sweden

Introduction

A system which can automatically recognize prosodic information in speech can be beneficial to a larger phonetically based speech recognition system. Information concerning stress, accent, phonetic focus and boundary signals can reduce lexical access time and provide information concerning phrase boundaries and syntactic structure (3, 8, 11). This paper represents a report from an ongoing joint research project shared by the Phonetics Departments at the Universities of Lund (Bruce, Eriksson and House) and Stockholm (Lacerda). The project, "Prosodic Parsing for Swedish Speech Recognition", is sponsored by the National Swedish Board for Technical Development and is part of the National Swedish Speech Recognition Effort in Speech Technology.

The primary goal of the project is to develop a method for extracting relevant prosodic information from a speech signal. Some issues relating to this goal are 1) What criteria can we use to recognize prosodic categories, 2) What kind of acoustic invariance relates to prosodic categories, and 3) What degree of success can we achieve in recognizing prosodic categories. Futhermore, by using a recognition approach to prosody, we hope to reach a better understanding of the mechanisms involved in human perception of prosody.

Our objective is to devise a system which from a speech signal input will provide us with a transcription showing syllabification of the utterance, categorization of the syllables into STRESSED and UNSTRESSED, categorization of the stressed syllables into WORD ACCENTS (ACUTE and GRAVE) and categorization of the word accents into FOCAL and NON-FOCAL accents. We also hope to be able to identify JUNCTURE (connective and boundary signals for phrases). We are currently working with 20 prosodically varied sentences spoken by two speakers of Stockholm Swedish.

Swedish is particularly interesting in terms of prosody recognition since the primary stressed syllable is characterized by having one of two tonal accents: Acute (Accent I) or Grave (Accent II). Identification of ACUTE accent can restrict and thereby facilitate the lexical search. Identification of GRAVE accent provides us with morphological information which can also facilitate lexical access. For example we know that the syllable following the stressed syllable of a grave accent word belongs to the same word (2). In languages without word-accent distinctions, however, our strategies for recognition of the categories STRESS, FOCUS and JUNCTURE may be applicable with modification.

The type and structure of the information to be presented to the recognizer has been based on a series of mingogram reading experiments (5). In the first experiment an expert in Swedish prosody (Gösta Bruce) was presented with mingogram representations of ten, phonetically balanced, unknown sentences showing a duplex oscillogram, fundamental frequency contour and intensity curve. On the basis of this visual information alone, he was able to identify 85% of all occurrences of the prosodic categories referred to above.

Descriptive rules were then formulated in which visually apparent acoustic criteria were used to describe the different prosodic categories. For example, a stressed syllable with a grave accent is described as having a falling fundamental frequency of a certain steepness and range where the beginning of the fall is synchronized with the vowel onset. A stressed syllable with an acute focal accent is described as having a rising fundamental frequency where the beginning of the rise is synchronized with the vowel onset. Crucial to this rule system is the relationship of fundamental frequency highs and lows between successive syllables. This relationship reflects the domain of accentuation. Also crucial to the rule system is the synchronization of fundamental frequency movement with vowel onsets. These two points

will be discussed later in terms of rule implementation.

In the second mingogram reading experiment, the descriptive rules were tested using non-expert mingogram readers. By applying the rules to mingograms of ten unknown sentences, two readers with no previous experience in visually identifying prosodic categories were able to attain scores of 78% and 69%. Our aim was to obtain similar or better results by implementing the rules on a computer system.

Our system for automatic prosodic recognition is comprised of three main steps (see figure 1). First, intensity and fundamental frequency are extracted from the digitized speech signal. Second, intensity relationships and fundamental frequency information are used to automatically segment the utterance into "tonal segments" which ideally correspond to syllabic units. Finally prosody recognition rules are applied to the tonal segments giving us prosodic categories as the output of the system.

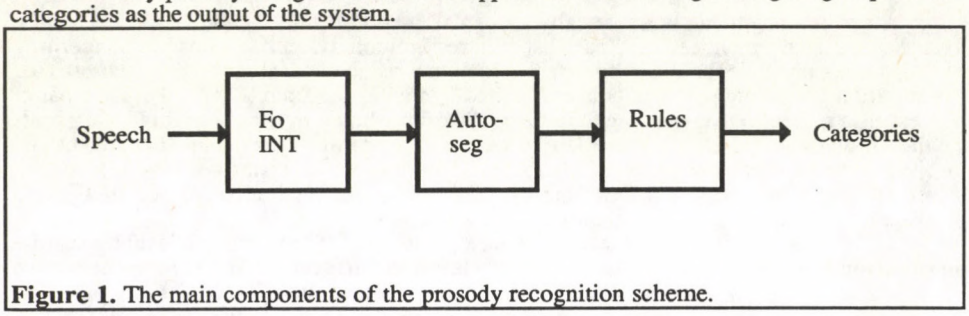


Figure 1. The main components of the prosody recognition scheme.

Automatic segmentation

A correct segmentation of the speech signal into syllabic units is of primary importance to the recognition system since the prosodic categories we are using are based on the syllable as the fundamental unit. It is also important that the system marks vowel onsets since vowel onsets make up the crucial synchronization points for identifying the prosodic categories using fundamental frequency movement.

The segmentation component has been designed using intensity measurements in much the same way as that described by Mertens (9). Similar algorithms have been described by Mermelstein (10), Lea (8) and Blomberg and Elenius (1). For a complete description of our algorithm see (6 and 7).

In short, the algorithm uses relationships between maximum and minimum values of both filtered and unfiltered intensity curves to make a broad segmentation. A -3dB threshold before the intensity maximum of each segment is used to locate the onset of the vowel for each syllabic nucleus. The end of the tonal segment is marked at the point where voicing ends prior to the next vowel onset, or if voicing continues, the end of the tonal segment will coincide with the next vowel onset. These tonal segments comprise the basic syllabic units for prosodic recognition. Maximum segmentation performance to date is 88%.

Rule implementation

Our preliminary strategy has been to reduce the information available to the recognizer in an attempt to attain the best results with the least possible amount of information. In this way we hope to isolate the most salient cues and build upon them to improve our results. It is clear from our descriptive rule testing that fundamental frequency information is crucial to the recognition of prosodic categories, especially word and focal accents. In our rule system Fo information is mainly expressed as relationships in Fo between successive syllables as this reflects the domain of accentuation.

Our task, then, is to reduce the analyzed Fo contour to a few values while maintaining critical information for recognition of prosodic categories. Evidence from our rule testing

indicated that an important area of Fo information is the average Fo level during the first 30-50 ms after vowel onset. This also corresponds to results from speech perception experiments (4). Another important area of information in the rules is the syllable final Fo level. We therefore decided to assign two Fo values to each tonal segment, average Fo during the first 30 ms (B) and average Fo during the last 30 ms of each tonal segment (E). This amounted to a linear stylization of the tonal contour. In order to test this stylization and see how much prosodic information is lost, we synthesized both speakers' productions of ten sentences using LPC synthesis with the stylized tonal contour as the pitch parameter. In several informal listening tests, the majority of the stylized sentences could not be distinguished from their original counterparts on the basis of intonation alone. Although the reductions did give rise to a few cases of clearly audible tonal deviations, the overall results give further strength to our preliminary method of reducing Fo information.

To incorporate Fo relationships between tonal segments, each segment is assigned two additional Fo values representing the high (H) and low (L) from the preceding (stylized) segment. Finally, two more values are assigned to each segment representing amount of (stylized) Fo change (C) during the segment and total duration (T) of the tonal segment.

In a first implementation of the rules using these six values, conditions for three word-accent categories (grave, acute+focal and acute+non-focal) were formulated based on the descriptive rules and on actual measurements of these values from the categories in question in ten test sentences. The conditions are listed in table 1.

Rule conditions for three word-accent categories. **Table 1**

Grave	**Acute+focal**	**Acute+non-focal**
C≤-20 Hz	C>5 Hz	-30 Hz<C<0 Hz
T>150 ms	T>100 ms	T>80 ms
B≥H-5 Hz	B>L-5 Hz	B<H
E<L-5 Hz	E≥H	E<L
	(B+E)/2>(H+L)/2	B<(H+L)/2

Where B=Fo beginning, E=Fo end, C=Fo change, T=duration of tonal segment, H=Fo high in preceding tonal segment, L= Fo low in preceding tonal segment.

A recognition routine checks each condition against the six values for each tonal segment. For each true condition, the segment receives one point for the category containing the condition. When all conditions are checked, the category having the most points is assigned to the segment. If two or more categories receive the same score, the following rule hierarchy applies: grave, acute+focal, acute+non-focal.

Finally a relative score threshold can be set where if the highest relative score does not reach the threshold, the syllable is assigned the category UNSTRESSED. If the score reaches the threshold, the category STRESSED is assigned by implication. For example with the threshold set at 0.75 (the value we are currently using) if grave receives two points, acute+focal three and acute+non-focal three, the segment will be assigned unstressed.

Results and discussion

The rule conditions for the three prosodic categories gave the following results when applied to ten test sentences: GRAVE 12 recognized of 13 occurrences, ACUTE+FOCAL 11 of 13, ACUTE+NON-FOCAL 7 of 10 and STRESSED 34 of 37. The category UNSTRESSED, however, was only recognized in 39 cases of 82 occurrences. These results combined make 103 recognized of 155 occurrences for a total score of 66%. This falls short of the scores for both our expert and non-expert mingogram readers.

The major problem with the rules seems to be an oversensitivity to Fo movement causing over half of the unstressed syllables to be categorized as stressed, in the majority of cases as ACUTE+NON-FOCAL. To a certain extent, this reflects the results of the expert

reader who identified 100% of the stressed syllables but only 73% of the unstressed (5). This problem can also be seen as an indication of the absense of a well-defined category boundary between the categories STRESSED and UNSTRESSED particularly in read speech.

In an attempt to improve identification of the unstressed syllables an integrated vowel intensity is measured for each tonal segment. A threshold is then set to separate the stressed from the unstressed vowels. At present, this threshold is set toward the unstressed end of the stressed-unstressed continuum, i.e. the threshold should exclude only unstressed vowels while letting some unstressed and all stressed vowels through. This threshold is applied to the tonal segments before the rule conditions are applied. Using the same ten test sentences, recognition of the UNSTRESSED category improved from 39 of 82 to 54 of 82 with only one STRESSED category being changed to UNSTRESSED by the intensity threshold. This improved the overall results from 66% to 77%, a score which almost equals our best non-expert reader.

Although the addition of other categories such as juncture and the problems involved in separating these cues from those of word accent may necessitate the use of additional parameter values for each tonal segment, our strategy of reduced information and stylization of the tonal contour seems to be a promising means of achieving prosodic recognition. A further sharpening of the rules and testing on a larger set of speech material should lead to improved results and increase our understanding of prosody in a speech recognition setting.

References

1. BLOMBERG, M.--ELENIUS, K. : Automatic time alignment of speech with a phonetic transcription. Proceedings of the French Swedish Seminar on Speech, eds. B. Guerin and R. Carré. Grenoble, 1985, 357--366.
2. BRUCE, G. : Structure and functions of prosody. Proceedings of the French Swedish Seminar on Speech, eds. B. Guerin and R. Carré. Grenoble, 1985, 549--559.
3. GIBBON, D.--BRAUN, G. : The PSI/PHI architecture for prosodic parsing. Proc. of the 12th International Conference on Computational Linguistics, ed. D. Vargha. Budapest, 1988, 1:202--204.
4. HOUSE, D. : Perception of tonal patterns in speech: implications for models of speech perception. Proc. of the Eleventh International Congress of Phonetic Sciences, ed. Ü. Viks. Academy of Sciences of the Estonian S.S.R. Tallinn, 1987, 1:76--79.
5. HOUSE, D.--BRUCE, G.--LACERDA, F.--LINDBLOM, B. : Automatic Prosodic Analysis for Swedish Speech Recognition. Proc. European Conference on Speech Technology, eds. J. Laver and M. A. Jack. Edinburgh, 1987, 1:215-218.
6. HOUSE, D.--BRUCE, G.--ERIKSSON, L,--LACERDA, F. : Recognition of prosodic categories in Swedish: Rule implementation. Working Papers 33, Department of Linguistics and Phonetics, Lund University. 1988, 161--169.
7. LACERDA, F.--BRUCE, G.--HOUSE, D.--ERIKSSON, L. : Prosodic parsing of Swedish, status report: Segmentation strategy. Proc. Seventh FASE Symposium, eds. W. A. Ainsworth and J. N. Holmes. Edinburgh, 1988, 1187--1195.
8. LEA, W. : Prosodic aids to speech recognition. Trends in Speech Recognition, ed. Wayne Lea, Englewood Cliffs, New Jersey: Prentice-Hall. 1980, 166--205.
9. MERTENS, P. : Automatic segmentation of speech into syllables. Proc. European Conference on Speech Technology, eds. J. Laver and M. A. Jack. Edinburgh, 1987, 2:9--12.
10. MERMELSTEIN, P. : Automatic segmentation of speech into syllabic units. Journal of the Acoustical Society of America 58 (4). 1975, 880--883.
11. VAISSIERE, J. : A suprasegmental component in a French speech recognition system: reducing the number of lexical hypotheses and detecting the main boundary. Recherches acoustiques CNET Lannion, 7. 1983, 111--112.

ROLE OF ACOUSTICS IN SEGMENTICAL TEXTUAL ANALYSIS

Klára KARIKÓ

Department of Hungarian Linguistics
Gyula Juhász Teachers' Training College, Szeged, Hungary

Textual analysis considering written language get the sentence boundaries delimated and further different structural, stilistic, sociolinguistic researches can be done. In an orally pronounced and later written text the commas deriding the sentence clauses can be grammatically explained but stating of sentence clozing punctuation mark depends on several factors (e.g. the interpunctuation of dialectal texts is mostly arbitrary and the text has to be arranged and formed because of its unconstructed message).

The aim of my research is to reveal those characteristic features of spoken language that can help to delineate the sentence boundaries of the analysed texts.

There are two closely connected approaches in my research work method:

a./ Considering the whole text I ascentain the smaller contextual and logical units (topic-comment units, types of differently constructed texts etc.)

b./ The registrated facts of speech dynamics (pause, intonation of a sentence, stress, rhythm etc.) are examined considering their mutual effect, their effect to the whole text, and the already revealed smaller units.

I would like to summarize the role of acoustics and dynamics in segmentical textual analysis on the basis of Hungarian texts.

The starting point in apprehension of acoustic of a spontaneous, free speech is that thinking and constructing of a text occur simultaneously. Acoustic elements appear in a free speech in various form and proportion.

The rough articulation, the interruptions, the continuity without a break, the broken units may be characteristic, and the dependence on parlance level and given situation as well.

Examining the intonation we always have to take into consideration the non-linguistic elements appearing in linguistic facts.

The semantics of intonation isn't a conceptual system, it only points out the socially valid stress that is unambiguous for the speaker and the listener.

Intonation is a kind of transition from natural, archaic phraseology to a better developed arbitrary system of notations that's why it has a supplementary role in communication.

The constant, socially valid intonation (question, imperative sentence, affirmative sentence) in concrete linguistic manifestation always appears as a characteristic feature of the speaking individual. Therefore it is very important to reveal and analyse these individual characteristics.

The most suitable approach in an intonation research is from methodological point of view to begin with the unit of form and content. This means that it is necessary to make a synthesis of the instrumentally measurable physical - physiological aspect of intonation and the linguistic - psychological one. So we can make a conclusion from form to function or we can interpret form from the differentitative, structuredistinctive role of intonation.

Intonation as a functional rhythmical unit bearing sense comprehends a definitive unit. The question is it's volume. Does it make dear the role of individual words in a sentence or does it comprise whole sentences and larger units by single rhythm motive?

To solve this question a survey of structural, modal and expressive function of intonation is a good help: a sentence in speech can be actually acceptable without grammatical form but it is unacceptable without intonation.

Intonation forms communicative units from words. This way it makes segments from speech (the speech units are delimited from each other) i.e. sentences can not be integrated in larger units.

It is more difficult to solve this question from the side of modal and expressive function of intonation. The subjective relation of the speaker to the "content units" is independent of its volume. Intonation: "comprehends large content units to wide range, demonstrates relation of individual elements in it, and gives contextual dissection of larger units.

The natural, inevitable connection of pitch and sensible dissection of communicative units predestinates intonation to this boundary-indicating role."

The above mentioned definition doesn't explain precisely the volume of "individual elements", larger units but referes to the relation of sentences coming one after the other. In speech course the different intonation forms often interpolate, dispose each other.

The search of smaller units, the question of disintegration leads to the boundary stating role of intonation.

The opinions on this question differ: the intonation can't be disintegrated; the real phrase, the period from one stress to another is the smallest intonation unit; there is a sentence clozing motive and the elements before it are irrelevant. The sentence closing motive of various languages is realized in different forms. In Hungarian language

sentence closing is mostly a falling intonation ending in pitch-note. The rising-"singing" intonation of the sentence ends is considered mistaken but the speech appearing mostly on television, radio with "correctly" falling, monotonious intonation is unnatural as well. In spoken language variance of the correct pronouncing norm bears specific function. It turns attention to the fact that it is not enough to make the acoustic analysis. We have to know the speaker's relation to the pronounced text, his intention with it and analysing the text one must make perfectly clear the role of intonation in communication.

The whole speech course is characterized by speech-pauses that usually depend on given situation. Pauses can denote the boundaries of linguistic-logical units of speech, instead of interjections or if missing they denote the connection between various elements. The main task of pause as a linguistic element is dissection. Pause appears in spontaneous speech course influenced not only by grammatical structure but a number of other facts as well. It can appear e.g. instead of conjunction or other grammatical element can stand instead of it. The sudden falling-rising intonation is suitable for making the feeling of pause. In subordinate sentences pause appears more frequently if the main clause isn't on the first place. Pause appears more rarely after and before inserted sentences in spoken language.

There are only a few examples studying the whole speech course and stating its acoustic elements. In constructing textual-phonetic system which validity overcomes the volume of a sentence, we have to take into consideration that it is not enough to summarize the acoustic facts but these facts bear informational, communicative surplus as well.

A phonetical sentence analysis can't be just a projection of tone, pitch, intonation, intensity, tempo, pause to the whole text. It can't be done so, because e.g. tone hasn't got special role in communication (but the changing of tone is very important in expressing feelings). Intonation that is always realized in sentence can't be called to account considering the whole text. The intensity specially in larger volume speech looses its function and can change causeless. The possibilities given by speech rhythm can be inconsistent as well. These above mentioned facts are unambigously parts of acoustics to have been taken in consideration but only if we counect them with the speaker's intention with the text and his relation to the textual composition.

Pitch, rhythm, intonation, stress, tempo and pause are characterising the whole text. In this sense it is suitable to use the concept of intonation to the acoustic-articulation complex which components are not always equal in speech-course. The role of intonation elements may change according

its relation to lexical units, grammatical facts, construction etc. This way intonation is a linguistic element not only from the phonetical but on functional side as well.

Revealing the types, constructions of intonation we can come to further facts in stating the orally pronounced sentence boundaries and this way not only functional but deeper contextual-formal sentence analyses are possible.

REFERENCES

1. Imre BÉKÉSI: Grammatics of thinking Bp. 1986.
2. Kálmán BOLLA: Intonational Research of Speech Course and Phonetic Delimation of Intonation Bp. 1979.
3. László DEME: Frequency Analysis of Sentence Construction Characteristics Bp. 1981.
4. Sándor KÁROLY: Sentence and Utterance Ethnography and Linguistics XXIV-XXV: 49-64
5. János PETŐFI: Text as an Interdisciplinal Research Medium Manuscript
6. Zoltán SZABÓ: New Methods in Textual Analysis Bucurest
7. Tamás SZENDE: Fundamental Facts of Speech Course Bp. 1976
8. Imre WACHA: Main Stylistic Categories of Spoken Language General Linguistic Studies X. 1974.

DIE GRAMMATISCHEN UND SATZPHONETISCHEN EIGENSCHAFTEN DER PARENTHESEN

Borbála KESZLER
Lehrstuhl für Ungarisch der Gegenwart
Loránd-Eötvös-Universität, Budapest, Ungarn

Obwohl die Definition der Parenthesen sowie der Bereich der Teile, die diese Funktion erfüllen, in der internationalen Literatur Gegenstand vieler Diskussionen sind, werden Existenz und Existenzberechtigung dieser Konstruktionen nie in Abrede gestellt. Für eine systematische Untersuchung des Problems findet man jedoch auch in der ausländischen Literatur nur äu3erst selten Beispiele und in der ungarischen schon überhaupt nicht. Über Parenthesen ist in der ungarischen Literatur nur in den Fachbüchern für Rechtschreibung und Sprachpflege sowie in jenen Kapiteln der Grammatiken zu lesen, die die keine selbständige Klanggestalt besitzenden Mittel der Satzkonstruktionen behandeln. Die Regeln der ungarischen Rechtschreibung (1) besagt zum Beispiel: "Wörter oder Wortgruppen, die man mit Absicht der Einschaltung (hervorgehoben von mir - B.K.) in den Satz einschiebt, werden in Kommata, Gedankenstriche oder Klammern eingeschlossen". Und in bezug auf die satzphonetischen Eigenschaften der Parenthese wurde schon vor langem festgestellt (2, 3, 4, 5), da3 für die Einschaltung gedrückter Tonfall, tiefere Tonlage, schnelleres Sprechtempo sowie Pausen vor und nach der Einschaltung bezeichnend sind.

Darüber hinaus gibt es aber noch recht viele andere ungeklärte Probleme im Zusammenhang mit der Paranthese. Solche sind zum Beispiel die folgenden:

1. Was charakterisiert die Parenthese vom grammatischen Gesichtspunkt aus? Können denn alle Satzglieder, Syntagmen oder Gliedsätze eingeschaltet werden oder gibt es dabei Verbotsregeln?

2. Hängt es allein von der Absicht des Verfassers ab, was er einschaltet, oder hat der Satz solche Teile, die sich automatisch hervorheben?

3. Können der gedrückte Tonfall, die tiefere Tonlage, das schnellere Sprechtempo und die Pausen vor und nach der Einschaltung mit konkreten Messungen bestätigt werden? Wie können sich diese Eigenschaften der Parenthese miteinander kombinieren und was sind ihre charakteristischen Kennzeichen?

1. In der ungarischen Literatur wurde bisher das grammatische Antlitz der Parenthesen nicht untersucht, obwohl eine solche Untersuchung äu3erst aufschlu3reich ist.

Die Parenthesen bilden keine einheitliche Gruppe, einerseits deshalb nicht, weil sie verschiedene Sprachebenen vertreten können: sie können nämlich Lexeme, Syntagmen, Sätze, mitunter sogar ganze Absätze sein, andererseits deshalb nicht, weil ein Teil von ihnen sich organisch (also auch grammatisch)

an andere Teile des Satzes knüpft (und eine erläuternde Bemerkung, einen zusätzlich angefügten Nachtrag, eventuell eine Hervorhebung ausdrückt), d.h. meistens die Hauptaussage präzisiert. Ein anderer Teil von ihnen knüpft sich grammatisch nicht, sondern nur semantisch an den Satz (oder eventuell an den Text). Es gibt schlie3lich auch Fälle, in denen die Parenthese nur den Wert eines modifizierenden Satzabschnittes besitzt.

1.1. Die Einschiebbarkeit von Wortgruppen, Wortgruppenteilen und Gliedsätzen, die organische Teile des Satzbaus sind, hängt mit der grammatischen und semantischen Selbständigkeit, der ergänzenden Bedeutung zusammen. Daraus folgt, da3 die Einschaltung der Hauptteile des Satzes sowie der rektionsartigen Ergänzungen nie möglich ist. Der Verbot gilt für diese auch, wenn sie durch einen Gliedsatz ausgedrückt werden. Infolge des ergänzenden Charakters können dagegen die nebengeordneten Teile, die freien Adverbialbestimmungen und die adjektivischen sowie appositionellen Attribute unbeschränkt eingeschaltet werden. Die aufgezählten Satzglieder lassen sich natürlich auch dann einschalten, wenn sie durch einen Gliedsatz ausgedrückt werden.

1.2. Eine andere Gruppe der Parenthesen knüpft sich grammatisch nicht, nur semantisch an den Satz (oder eventuell an den Text). In diesen Fällen ist die Parenthese oft ein Mittel der Doppelmitteilung. Der anorganische, "aus einer anderen Sprachsphäre stammende Teil", wie er in der Grammatik von Quirk--Greenbaum (6) genannt wird, kann die beiläufigen Bemerkungen des Autors bzw. des Verfassers des Textes ausdrücken oder als Mittel der Hinwendung zum Publikum dienen, wobei der Hauptteil die Handlung weiterführt. Fónagy (7) erwähnt, da3 die Klammern bei Proust (die Klammern sind, wie bekannt, die ausdrucksvollsten Mittel der Einschaltung) Leo Spitzer an runde Fensterchen erinnerten, aus denen sich der Schriftsteller seinem Leser vertraulich zuneigt. Von manchen werden auch die eingeschobenen Anreden für Parenthesen gehalten (8).

1.3. Schlie3lich gibt es auch Fälle, in denen die Parenthese lediglich die Meinung, die Stellungnahme des Verfassers des Textes zum Satz enthält. In diesen Fällen kann häufig nicht einmal über eine semantische Beziehung gesprochen werden, die Einschaltung hat also nur den Wert eines modifizierenden Satzabschnittes und dieser modale Wert kann manchmal auch pragmatische Komponenten haben.

Es ist äu3erst wichtig, bei den Parenthesen zu wissen, da3 für sie, so unabhängig sie auch manchmal vom Satz grammatisch sind, immer eine kommunikative Unselbständigkeit, die Zugehörigkeit zu einem konkreten Satz oder eventuell zu einem konkreten Text bezeichnend ist.

2. Um zu beweisen, da3 die Parenthese in manchen Fällen nicht blo3 von der Absicht des Sprechers abhängt, habe ich eine Satzreihe aus 10 Sätzen zusammengestellt, in der, meiner Meinung nach, sich gewisse Teile der Sätze automatisch hervorheben. In den Sätzen machte ich den Platz der Einschaltungen durch keine Satzzeichen erkennbar, den Vorlesern habe ich aber gesagt, da3 die Satzzeichen im Satzinneren nicht gesetzt sind. Über die Sätze machte ich mit Hilfe von 4 Vorlesern (mit je 2 Männern

und Frauen) Tonaufzeichnungen. Dann wurden das Oszillogramm, die Intensitäts- und die Tonfallkurve der Sätze im experimentalphysischen Laboratorium des Insitutes für Sprachwissenschaft der Ungarischen Akademie der Wissenschaften mit Me3instrumenten (Tonhöhenmesser Typ FFM 650, Intensitätsmesser Typ IM 360, Mingograph Typ EM 34 T) produziert, die zur Untersuchung der sog. satzphonetischen (suprasegmentalen) Eigenschaften dienen. Die Integrationszeit der Analysen stellten wir auf 10 ms ein. Die drei Registraturen (die oszillographische Schwingungsform, den Melodieverlauf und die Intensitätskurve) konnte ich synchron auf dem mit einer Geschwindigkeit von 100 mm/s hergestellten Diagramm studieren. Das Experiment hat bestätigt, da3 sich bestimmte Teile mancher Sätze tatsächlich automatisch hervorheben und eindeutig die bezeichnenden Eigenschaften der Parenthese aufweisen. Solche Teile sind z.B. bestimmte Typen der appositionellen Attribute, das zweite Glied gewisser nebenordnender Konstruktionen, manche Adverbialbestimmungen usw.

3. Zum Schlu3 habe ich aufgrund einer aus 10 Sätzen bestehenden Satzreihe auf die oben beschriebene Weise (wobei die Satzzeichen im Satzinneren nunmehr gesetzt waren) die Dauer (in ms) der Pausen vor und nach der Einschaltung, die Verhältnisse der Tonhöhenbewegungen, des Sprechtempos sowie der Intensitätswerte der Parenthesen untersucht, indem ich sie mit den Tonhöhenbewegungen, dem Sprechtempo sowie den Intensitätswerten der Teile vor und nach der Parenthese verglichen habe.

Das gleichzeitige Erscheinen der vier charakteristischen Eigenschaften der Parenthese habe ich natürlich nicht erwartet, denn sie sind zwar üblich, jedoch nicht unbedingt notwendig. Deme zufolge (2) "tragen die zusammen erscheinenden vier satzphonetischen Eigenschaften nicht alle Hauptfunktionen, so können sie sich in begründeten Fällen voneinander trennen".

Die instrumentelle Untersuchung ergab folgendes:

Die charakteristischste und allgemeinste Eigenschaft der Parenthese sind die Pausen vor und nach ihr. Sie sind in 90% der Fälle zu finden. Die Dauer der Pausen hängt von den grammatischen Besonderheiten und der Länge der Parenthese ab, kann aber Unterschiede in ein und demselben Satz auch je nach dem Interpreten aufweisen und auch dadurch bestimmt sein, in welche Satzzeichen die Parenthese im geschriebenen Satz eingeschlossen wird. Die Pause vor dem Einschub ist meistens kürzer als die nach ihm. Meinen Untersuchungen nach macht der Durchschnittsunterschied etwa 50 msec aus.

Die Parenthese stellt meistens eine selbständige Melodieperiode dar, wenngleich sie kein ausgezeichnetes Melodiemodell hat. In 75% der Fälle ist ein starker Tonfall am Anfang der Parenthese zu beobachten. Eine Ausnahme bilden dabei die Parenthesen, die eine partielle Apposition, eine einschränkende Adverbialbestimmung oder eine begrenzende Fügung enthalten.

In der Fachliteratur wird oft erwähnt (s. oben), da3 auch das schnellere Sprechtempo für die Parenthese charakteristisch ist. Diese Hypothese wurde aber bisher mit instrumenteller Untersuchung an ungarischem sprachlichem Material noch von niemand bestätigt. In 60% der Fälle konnte ich eine Tempoerhöhung tatsächlich beobachten. Sie machte mitunter sogar einen Unter-

schied von 2-3 Laut/sec aus. In 50% der Fälle änderte sich jedoch das Sprechtempo nicht oder es wurde sogar langsamer. Letzteres kam hauptsächlich vor, wenn die Parenthese keine Erklärung, beiläufige Ergänzung oder Bemerkung, sondern gerade eine hervorzuhebende Aussage enthielt.

In bezug auf die Intensitätswerte der Parenthese habe ich die Erfahrung gemacht, da3 die Parenthese in 50% der Fälle von geringerer Intensität ist als ihr Umfeld. Am Anfang der Parenthese kann der Intensitätsrückfall bis von 5-8 dB sein. Die Intensitätswerte gewisser Parenthesen (wie zum Beispiel der partiellen Appositionen, der einschränkenden Adverbialbestimmungen usw.) sind jedoch höher als die ihres Umfeldes.

Der Umfang des bearbeiteten Materials lä3t natürlich keine allgemeingültigen Schlu3folgerungen zu, die Ergebnisse und Lehren beleuchten jedoch, in welche Richtung die Forschungen weitergeführt werden sollen. Zunächst müssen die Untersuchungen auch auf spontanes sprachliches Material ausgedehnt werden und dann müssen die suprasegmentalen Eigenschaften der Parenthesen den einzelnen grammatischen Typen nach untersucht werden, wobei auch die Voraussetzungen und Regeln der automatischen Hervorhebung zu beachten sind.

Literatur:

1. A magyar helyesírás szabályai (11. kiadás) /Regeln der ungarischen Rechtschreibung (11. Auflage)/. Akadémiai Kiadó, Budapest, 1984.
2. DEME, L.: A kiejtés törvényeinek tanítása /Unterricht der Gesetze der Aussprache/. Magyar Nyelvőr 94, 1970, 270-280.
3. DEME, L.: A helyes magyar kiejtés kérdése /Die Frage der richtigen ungarischen Aussprache/. In: Nyelvművelésünk főbb kérdései /Wichtigere Fragen der ungarischen Sprachpflege/. Akadémiai Kiadó, Budapest, 1953, 199-239.
4. DEME, L.: Helyesírási rendszerünk logikája /Die Logik des ungarischen Rechtschreibesystems/. MNyTK. Nr. 110, 1965, 32.
5. PÉCHY, B.: Beszélni nehéz! /Sprechen ist schwer!/. Magvető Könyvkiadó, Budapest, 1974, 167-168.
6. QUIRK, R.--GREENBAUM, S.: A University Grammar of English. London, 1977, 459.
7. FÓNAGY, I.: Írásjel (címszó) /Satzzeichen (Stichwort)/. In: Világirodalmi Lexikon 5. Akadémiai Kiadó, Budapest, 1977, 111-123.
8. BU3MANN, H.: Lexikon der Sprachwissenschaft. Stuttgart, 1983.

VOLUMINOUSNESS AS A FEATURE OF THE ARTICULATORY BASE OF LANGUAGE

KLIMOV N.
Department of German Phonetics of the
Maurice Thorez Moscow State Institute
of Foreign Languages

According to the theory of movement formation by N.A. Bernstein (1) any complex activity (as well as the speech organs activity) is regulated on several neurophysiological levels. In the framework of the hierarchy of levels there exists a level ('the synergic level') checking the realization of more general and relatively permanent movement features. A person's gait, handwriting, as well as language peculiarities of the phonetic base may serve as an example of these features.

In the description, didactically orientated Phonetics the notion of the phonetic base as a unity of the most indexical and permanent features of the acoustic form of the language has been used since the XIXth century. However, up to the present day the outline of these features is of an impressionistic character as a rule. We can only trace out some few attempts of experimental acoustic studies of these features (2,3).

An object of the present study is one of the indexical features of the articulatory base, comparative to the German and Russian languages. The study is based on a supposition that certain languages have a timbre colouring of their own. The presence of a specific timbre colouring is reflected in our mind in the usage of corresponding epithets (ex: glottal, hoarse, etc.). If the existence of a specific general language timbre can be diagnosed then the question of its physiological interpretation, i.e. the description of that form of the mouth resonator that can be regarded as a source of a particular spectrum picture (the acoustic perception), appears to be relevant.

A long-term-spectrum which is frequently used in the Acoustical Phonetics, and in the contrastive aspect, in particular, may serve as an adequate instrument of language timbre (4,5,6). The analysis of a number of studies where the long-term-spectrum was used shows that the latter is a function of several variable quantities. The most indexical of them are individual speech characteristics, the segmental structure of the text under investigation, the intensity of a speech signal and specific language feature is less relevant for the overall picture of spectrum than the other factors. Consequently, the correct application of the long-term-spectrum for the investigation of the specific language features calls for a thourough control of other factors referred to above.

To neutralize the influence of individual speech pecularities we employed speakers-bilinguals who have an equally good command of German and Russian. To reduce the effect of the factor 'the segmental structure of the text' we chose texts of relatively long duration (about 60 sec. each). Finally, the intensity of spectrum was normalized, i.e. the intensity data of each frequency range were not taken independently but in their relation to the long-term-spectrum intensity of the text.

15 speakers-bilinguals (all male) participated in the experiment. Each of the subjects was to read out 4 narrative texts (each 60 sec. of duration): 2 texts in German and 2 in Russian. The subjects' production was transformed into graphic records of the long-term-spectra using two spectrographs 'IS-1' (an apparatus designed at the Phonetic Laboratory of the Maurice Thorez Moscow State Institute of Foreign Languages) and Bruel and Kjaer 2131. A frequency range from 88 to 4000Hz was tested. The results drawn

from the analysis of the long-term-spectra graphic records of the German texts demonstrated that 13 out of 15 speakers showed an increase of intensity (about 5 dB) within the frequency range between 700 and 1400 Hz.

A physiological approach to the acoustic data obtained was also attempted. 10 auditors -- native speakers of Russian -- with a good ear for timbre and imitation abilities were to listen to short texts in Russian and German and define qualitative pecularities of German timbre. The following assumptions arose from the questioning of auditors:

1) the phonation of German speech can be characterized as more "volumetric";

2) widening of the mouth resonator resulting from the raising of the soft palate and lowering of the tongue body should be regarded as a physiological basis of voluminousness;

3) the German vowels /a:/, /a/, / / as well as diphthongs /ao/, /ae/ show the largest voluminousness of phonation.

It should be supposed that the usage of these very sounds in speech (their frequence of occurrence is 61% of the total number of German vowels) results in the acoustic perception of the entire colouring of speech.

To test the hypothesis of intensity spectrum increase in the frequency range between 700 and 1400 Hz with widening of the back part of the mouth resonator the so-called "matched--guises" were used.. An experienced phonetician was to read out two groups of, syllables like CV, including the vowel sounds / /, / /, where one group was to be pronounced with an emphasized "voluminousness", and the other - neutrally. The comparison of the integral spectra obtained for each group showed that the "volumetric" variant was characterised by a sharp (3-4 fold) increase of intensity in the frequency range between 700 and 1400 Hz.

Evidently, a further progress in experimental contrastive studies in the above described field may contribute to a certain extent to the creation of systems that will ensure an automatic recognition of the language, as well as more sophisticated visual aids used in teaching foreign Phonetics.

REFERENCES

1. N.A. BERNSTEIN: K voprosu o prirode i dinamike koordinatsyonnoi funktsyi. (Uchyonye sapiski MGU. Vyp. 90, 1945).
2. F. NOLAN) The phonetic bases of speaker recognition. Cambridge, 1983.
3. I. LAVER: The phonetic description of voice quality. Cambridge, 1980.
4. T. TARNÓCZY: Das durchschnittliche Energie-Spectrum der Sprache (Acustica. Vol. 24, H.2, 1971. pp. 57--74).
5. I. ZALEWSKI--W. MAJEWSKI: Polish speech spectrum obtained from superimposed samples and its comparison with spectra of other languages. (7-th IC on Acoustics. Budapest, 1971, 24 C.22, p.249--252).
6. B. HARMEGNIES--A. LANDEREY: Language Features in the Long-Term Average Spectrum. (Revue de Phon. Appl. 1985, 73/75 p. 70--80).

ON RESEARCHING THE PITCH OF SPEECH BASED ON READ LITERATURE TEXTS

Éva KINCSES KOVÁCS
Institute of Literature, Hungarian Academy of Sciences, Budapest

Place of the subject

Investigating the latest trends in the 20th century linguistics it is the text and the supersegmental and extra-linguistic elements that primarily attract our attention. Recently the way of look at language has changed, the former centres of interest have shifted from the historic approach first to synchronic research and then to deep structures. We must go back almost 100 years in the giwen special literature/bibliography, if we want to investigate the treatment and description of the supersegmental elements in the last quarter of the 20th century. It isnot the distance in time, but the interdependence of the various branches of science (music, literature, aesthetics, physics, pshychology, physiology, sociology, etc.) that makes the survey more difficult. That is why linguistic, text-phonological and instrumental phonetic methods should be applied primarily in the research, but adopting the methods of musicology and psychology as well. The present experiment was carried out with the demand of interdisciplinarity and exactitude.

The sciences of languages and literature have been able to develop so rapidly owing to first of all the flow of communication and information: the improving and widespreading of the technical instruments such as television, radio, tape recorder, cassette player, video and film laid behind all this. The science of our days has established the closest ever contact with real practice on the one hand, and the masses of recipients and appliers on the other.

The literary text as the subject of analysis

Independently of written literature, the presentation and reception of literature -i.e. the developing and transmission of ´poetic message´- were the parts of human culture in the past and also are in the present. We have new vistas to admit and analyse pieces of arts, including literature, by way of developing the human organs of sense and the nervous system, and parallelly, by bringing up to date the instruments invented by humans. Textology as an individual branch of knowledge -as far as its methods and definition are concerned- is a field, crystallizing nowadays. ´It became a true integrating discipline..., has claim to linguistic and literary branches of science, fuses with and draws a parallel between such disciplines as stylistics, poetics, and theory of literature´ (Zoltán Szabó: The new Ways of Textual Analysis. 1982. Kriterion. 175-6).

László Deme investigated text as an announcement (communication) in his essay called 'Textity and some characteristic features of text-cohesion (In: Essays on the textology on the contemporary Hungarian language. Ed. E. Rácz and I. Szathmári. 1983. Budapest). He turned the attention especially to the inner construction of a text, its units and the coherence of the content included in them. He dealt with paragraphs, chapters and even the structure of the whole text when dividing the text. His basic category and fundamental unit was sentence, the smallest link-like unit of speech. The essence of Deme's functional-constructional definition of the text is that the text plays some role and it is in a particular position. These two factors together determine its length, form and formulatedness taken as a function of the speaker, the listener and the text (reality). The 'textity of a text' depends not on to what extent it is expounded and formulated. One single defective sentence can as well have text-value on the basis of common language, preliminary knowledge and common precedents.

Each of the 14 pieces of literature chosen for analysis suits the criterions of the textity-description above. Their quality as texts is unquestionable. But the degree of their formulatedness/or rather the organization of text according to my usage of terminology/ is different. The read literary texts are different as far as their types are concerned, but they are roughly the same considering their length and duration. The forms, variants and ways of appearance resulting from the degree of the definiteness of the text are as follows: poem, prose, free-verse, transitional categories. We cannot deal with the precise systematization of the texts in verre here (such as accentual and metrical versification).

Course of the analysis

Not each of the written texts can be enjoyale and well interpretable when they are performed. József Bakos with his modern experiments searched for an answer to the question why and how an announcement becomes a written or an oral text in his essay called Creation, interpretation and making the texts oral which are for reading and performing (In: Essays...6.) Certain texts can be impressive when reading aloud , though originally they were not devoted to that. Each of the artistic texts can be changed into an oral one and is suitable to be performed. In most of the cases they were devoted even to this. A poem remains dead material without sounding it. Similarly a prose-work, and especially a free-verse and the different transitional categories hide their inner constructions, their rhythm. Only reading aloud can result in an adequate analysis and interpretation. Summing up the whole idea we may say that 'silent reading is the death of literary texts'. In the case of a longer prose-work of course this approach is practically impossible,

but such interpretation of the most characteristic parts is neither useless nor negligible. For example not only the poems of Petöfi are highly rhythmical, but his prose as well.

Method

We investigated female and male voices (read literary texts) at the Institutes of Literature and Language of the Hungarian Academy of Sciences. Members of the Department of Phonetics (under the direction of Gábor Olaszy and Péter Nikléczy, and on the basis of András Kecskés's previous research) prepared recordings from read literary texts with an instrument called the Mingograph 34 T Elema Schönander (Stockholm). Each reader performed/sounded her or his text three times. The texts were of different types but approximately the same regarding their length and duration. The texts lend themselves for evaluation of both historical and synchronic nature. They give a cross-section of the work of our most distinguished poets from the 19th century until these days. This research can have and it really has some socio- and psycholinguistic aspects: the differences between female and male voices can be proved not only with physical parameters and physiological facts, but with loud reading of literary works as well. It is worth mentioning that men hardly or not at all volunteered for this seemingly easy task, in spite the fact that they often got near to microphon by virtue of their profession and qualification. Women read with pleasure and fairly well the classic artistic texts known from their previous studies, like the works of Kölcsey, Petöfi, Ady, Kosztolányi, Attila József, Csáth, Lörinc Szabó, Kassák, Utassy, Kányádi, Esterházy. The women's emotional attitude made the analysis of the recordings easier, the richness of affectiveness came to light/rose to the surface by investigating melody and pitch. The emotional impregnation, and the vibration are well provable with the help of different perceptive 'instruments' (hearing-both its direct way and the way made indirect by a tape recorder, melody pattern securing in a diagram, curved lines and statistics which can be described and made visible by function diagrams).

Role of pitch in human speech

Pitch, intonation or melody is the oldest manifestation of human speech, in which we can find the trace of music. This spontaneous musicality formed the basis of speech, later becoming articulated, of mankind, which was divided into nations and languages.
From among the two basic factors of language/speech intonation means the elements beyond the segments. Intonation

manifests itself in the changes of pitch, which in this waycan be recorded by instruments. Segments are the phonemes andthe sounds of speech, and the constructions of higher level, the latter being organized from the former ones. On the other hand the sound-features, the different sentence and text-phonological devices built on segments are called suprasegments. The speaker synthetizes the form possible or necessary for him or her from the available, above mentioned store. The synthesis is established from two interwoven threads revolved spirally by an imaginary vertical axis, which is a function of time and space. We speak about the process of organization of segments and suprasegments, in which pitch -being a true mirror of the given situation and emotional state-, plays a significant role in giving the final form. While the melodious nature of spontaneous speech is the 'unartful companion' of the statement with the words of Zsigmond László, the individual style of a prose-writer determines certain melodious forms. But the melody-constituent of a poem cannot be only a decorative element, it must contribute creatively to the essence of the poem (In: Zs. László: Rhythm and melody. Poetry and musicality. 1985. Budapest).

Pitch carries separate semantic function, a child feels and even produce its changes very early.

Pitch is an archetypal form characteristic of human speech, which has a distinguished role among the supersegments. It is because it possesses individual function and style creating power, carries meaning, and because it is one of the organizing forces of the sounded text. Pitch is a formal result of some grammatical-topical constructions and it is more or less defined in the different languages.

Investigating the recordings of my female and male readers it became possible to carry out paralelly both the simple perceptive (hearing, seeing) and the instrumental evalutaion. What provided occasionally an opportunity for contradictions subjectively (e.g. simple delivery, unrhythmical division of the text, unjustified changes in pitch, gabble, not clear intonation and voice-production, too long pauses, the possibility of misinterpretation resulting from imperfect readings) can be presented and corrected with the help of the instrumental recordings. From among the various individual reading an average curve is outlined, in which the similarities are dominant in spite of the differences and which give help to a new reader. It is writing out in score in a sense that goes on, but only sketchy, the contradictions cannot be standardized, they remain individual beside their common features.

EXPERIMENTAL STUDIES OF POETIC RHYTHM

Ilse LEHISTE
Department of Linguistics, Ohio State University, Columbus, OH 43210, U.S.A.

Introduction

The paper constitutes a progress report on a new project, currently under way, designed to analyze the metrical structure of orally produced poetry in various languages, using acoustic phonetic techniques. The motivation for the project arose from the following considerations.

Rhythm is an essential part of the suprasegmental structure of a language. It appears reasonable to look for rhythm where one can be sure rhythm can be found: in the metric structure of poetry developed in a given language over the years. It is a basic assumption in my current study that the suprasegmental system of a language is crystallized, as it were, in the metric structure of its traditional poetry. Patterns that may be imperfectly realized in prose may be manifested in a more regular fashion in poetry; the rhythmic structure of poetry may just represent what for the realization of segmental sounds has sometimes been called "maximally differentiated style". I submit that for the suprasegmental system of a language, poetry represents that maximally differentiated style. One may observe, on occasion, a "creative tension" between poetic form and the structures established by, e.g., word-formation rules; but this tension presupposes the existence of patterns and by its very systematicity can be used to deduce the patterns. An example might be found in the Finnish Kalevala verse, where short word-initial syllables are systematically excluded from ictus position, even though word-level stress always falls on the first syllable of a word.

While I am particularly interested in studying the phonetic realization of the metric structure of folk poetry, I expect to find it informative as well to study the adaptation of certain classic metres to a particular language.

I have carried out a number of preliminary, primarily descriptive studies of the realization of similar metres in different languages (Lehiste and Bond 1984; Lehiste 1984; Lehiste 1986; Lehiste 1987; Lehiste to appear), and am currently engaged in a more systematic study, of which the work reported in the present paper constitutes a relatively small part.

Methods

The present paper compares the realization of the trochaeic

and dactylic metres in poems produced orally by one speaker of Finnish and one speaker of Faroese. The Finnish poems were selected by Kalevi Wiik, Professor of Linguistics at Turku University in Finland, with the help of Lea Rojola, Assistant Professor for Finnish Literature at Turku University. The poems were recorded by seven speakers at the Phonetics Laboratory of Turku University Sept. 26-30, 1988, with the technical assistance of Lauri Kurki and Riku Kivinen.

The Finnish materials discussed in this paper consist of two trochaeic poems, "Vastavirtaan" by Juhani Siljo (32 lines) and "Kapina" by Lauri Viita (38 lines), and one dactylic poem, "Tanssilaulu", by Juhani Siljo (16 lines), read by Kalevi Wiik.

The Faroese poems were selected by Mr. Jogvan Isaksen, Adjunkt at the Arnamagnaean Institute of the University of Copenhagen, native speaker and lecturer of Faroese. Mr. Isaksen also served as informant. The recordings were made at the Phonetics Laboratory of the University of Copenhagen, with the help of Professor Jørgen Rischel, on July 11, 1988. The materials discussed in the current paper consist of the first ten stanzas each of two ballads, the popular ballad "Brynhildar táttur" and "Ormurin langi" by Jens Kristian Djurhuus, and the poem "Fast stóð í fonnum" by Rasmus Effersøe. The number of lines was 40 for "Brynhildar táttur", 40 for "Ormurin langi", and 24 for "Fast stóð í fonnum".

The tapes were processed at The Ohio State University by the MacAdios Waveform Analysis System (Spectrographic Analysis Program, GW Instruments), implemented on a Macintosh Plus computer. The signal was sampled at 10 kHz, and low-pass filtered at 4.7 kHz. Spectrograms were likewise made on a Voiceprint spectrograph. Measurements of the durations of metric feet were made by I.L. with an attempted precision of $\pm$ 0.5 mm, corresponding to $\pm$ 3 msec. While the averages in the tables are presented with a precision of 1 msec, no claim is made that the actual measurements could achieve that level of precision.

Results

Table 1 presents average durations, in milliseconds, of medial metric feet in the described trochaeic and dactylic poems. The Finnish trochaeic poems did, in fact, consist of disyllabic metric feet; the Finnish dactylic poem contained disyllabic feet as well. The Faroese ballads contained a majority of disyllabic feet, as did the third poem, even though it had been designated dactylic by the informant. This circumstance turned out to be fortuitous, since it makes it possible to compare disyllabic and trisyllabic metric feet within the same line. The first and last metric foot were left out of the calculations, the first because of the impossibility of measuring the duration of initial plosive consonants, and the last because in many instances the last foot was monosyllabic.

The boundaries between metric feet were established according to the following criteria. A single intervocalic

consonant was considered part of the syllable starting the following metric foot. In the case of intervocalic consonant clusters, the boundary was assumed to occur before the last consonant of the cluster. In the case of Finnish geminates, if there were no acoustic cues to the presence of the boundary, the boundary was assumed to occur at a point preceding the onset of the vowel by an amount corresponding to the average duration of a single initial consonant of the same class.

In Finnish disyllabic metric feet, the duration of the individual syllables was also established; in Faroese, the metric feet were not further subdivided into syllables.

Table 1: Average durations (in milliseconds) of medial metric feet in trochaeic and dactylic poems in Finnish and Faroese. S=syllables; N = number of occurrences; $\bar{x}$ = average; σ = standard deviation.

	Finnish			Faroese			
	Trochaeic	Dactylic		Trochaeic		Dactylic	
	(2-S)	(2-S)	(3-S)	(2-S)	(3-S)	(2-S)	(3-S)
N	158	15	33	104	31	147	55
$\bar{x}$	473	480	660	411	458	400	459
σ	110.1	88.9	155.6	71.3	78.2	71.5	85.4

The table reveals at least two interesting differences between the two sets of materials. In Finnish, the difference between disyllabic and trisyllabic metric feet is, on an average, 184 msec; In Faroese, this difference is only 53 msec. It appears questionable whether a difference of this magnitude - a difference of approximately 13% - is perceptible at all. (The question of just noticeable differences in the duration of one of four signals needs to be explored.) In Finnish, the difference might well amount to the duration of an added syllable, whereas in Faroese this does not seem to be the case.

A further difference between the two sets of data appears in the measures of variability: the standard deviations of the Finnish metric feet are larger in every case than the corresponding standard deviations in the durations of Faroese metric feet. A possible reason might be the fact that the duration of a metric foot in Finnish depends on the type of syllables constituting the foot. Table 2 presents average measured durations of the disyllabic metric feet analyzed in the present study, classified according to syllable types constituting the feet.

Table 2. Average durations (in milliseconds) of medial metric feet in two trochaeic Finnish poems. N = number of occurrences; $\bar{x}$ - average ; σ = standard deviation.

Foot type	Short-short	Short-long	Long-short	Long-long
N	41	9	81	27
$\bar{x}$	380	524	479	606
σ	75.1	104.3	92.3	99.7

Discussion

The table reveals, first of all, that the temporal structure of a disyllabic metric foot in Finnish depends on syllable duration rather than on word-level stress. Stress always falls on the first syllable of a word; in a trochaeic metric foot, ictus falls likewise on the first syllable. The durations of Short-long and Long-short metric feet indicate that the length of a long syllable is not reduced in unstressed position.

In a mora-counting analysis of Finnish quantity, a short syllable would contain one mora, and a long syllable two moras. The four metric foot types presented in Table 2 would exhibit a regular reduction of the duration of a mora: 190 msec for Short-short, 175 msec for Short-long, 160 msec for Long-short, and 150 msec for Long-long. The mora thus does not seem to constitute a regular unit of temporal programming, as it does in Japanese (cf. Lehiste, to appear).

It is also worth noticing that the difference between Finnish trisyllabic metric feet and the Long-long type disyllabic metric feet (54 msec) is comparable to the difference between trisyllabic and disyllabic metric feet in Faroese (53 msec). While it appears likely that a difference of 606 and 380 msec is perceptible, the differences between 660 and 606 (Finnish trisyllabic and Long-long disyllabic) on the one hand, and between 459 and 400 (Faroese trisyllabic and disyllabic) may not be. Perceptual tests are called for to establish the degree of isochronicity between these metric foot types.

References

1. LEHISTE, I.--BOND, D.: The phonetic realization of the trochaeic metre in Latvian and Estonian. Journal of Baltic Studies 15, 4. 1984, 293--302.
2. LEHISTE, I.: The metric structure of a recited Finnish spell. Ural-Altaische Jahrbücher 4. 1984, 83--89.
3. LEHISTE, I.: Phonetic realization of metrical structure in some types of Estonian verse. Ural-Altaische Jahrbücher 6. 1986, 11--20.
4. LEHISTE, I.: Rhythm in spoken sentences and read poetry. Phonologica 1984. Cambridge University Press, 1987. 165-173.
5. LEHISTE, I.: The phonetic realization of the haiku form in Estonian poetry, compared to Japanese. To appear.

THE DURATION OF DISYLLABICS IN THE SUONIKYLÄ DIALECT OF SKOLT SÁMI

Zita McROBBIE-UTASI
Linguistics Department
University of Manitoba, Winnipeg, Canada

In Skolt Sámi there is an optional rule that either reduces word-final short vowels or deletes them. The purpose of this report is to examine the effect of the reduction or the drop of the vowels in question upon the duration of the preceding consonant(s) (consonant centre) and the first syllabic vowel (vowel centre). Durational measurements of several hundred disyllabics indicate attestable durational increases in most such cases; the ratios of the vowel centre and the consonant centre, however, remain basically the same in these stress-group locations.

Previous analyses of Sámi quantity have clearly established that disyllabic units (stress-groups) are to be regarded as the domain of quantity (1),(2). The terms that have been used for referring to the main stress-group locations - vowel centre, consonant centre and latus - denote the fact that syllables have no relevance in quantity distributions, all structural restrictions being centred on syllable boundaries (3),(5). The test-words that have been examined for this paper are all disyllabics.

The recording of the Suonikyla disyllabics took place at the University of Manitoba. The native speaker of this Skolt Sami dialect was asked to place the test-words in a sentence frame cie'lǩ ... e'pet 'say ... again'. The recording was made with a Scully Full-Track Broadcast Machine tape recorder; the tape speed was 7.5" per second. The software employed for making durational measurements is the DSPS Digital Signal Processing Software, Real Time Signal Lab. This software is designed to produce spectrograms together with continuous wave-form displays. The program allows one to expand the spectral and wave-forms, facilitating the achieving of the desired accuracy. Simultaneous digitalization gives a read-out in milliseconds.

The material was organized as follows: all disyllabics measured were grouped into five main structural types. These structural types are the same as the ones that are referred to in E.Itkonen (2) as $\breve{3}\breve{x}\breve{x}^{\breve{3}}$, $\breve{3}\breve{x}\breve{y}^{\breve{3}}$, $\breve{3}\breve{x}\breve{3}$, $\breve{3}\breve{x}\breve{x}^{\breve{3}}$ and $\breve{3}\breve{x}\breve{y}^{\breve{3}}$. (The symbol $\breve{3}$ indicates a vowel, $\breve{x}$, $\breve{x}\breve{x}$ and $\breve{x}\breve{y}$ stand for single consonants, geminates and consonant clusters respectively). Only disyllabics with a non-contracted second syllabic vowel were considered for the present investigation, because the optional phonological rule in question does not apply to contracted vowels.

The five main structural types will be referred to as Type 1, Type 2, Type 3, Type 4 and Type 5. Each of these main types is subdivided into several groups according to the consonant centre. Type 1 and Type 2 have long consonants in the consonant centre (geminates and consonant clusters respectively), Types 3, 4 and 5 have short consonants in the consonant centre (simple consonants, geminates and consonant clusters respectively). A code for interpreting the consonant combinations in the five different structural types is provided in the Appendix.

Durational measurements were made of the vowel centre and the consonant centre, and also of the latus when it was present. Latic durations average 90 mscs. In the Tables that follows I summarize the analysis of the measurements obtained. Mean durations ($\bar{x}$) and standard deviations (SD) are given for the vowel centre and the consonant centre separately for disyllabics with and without a latic vowel. Similarly, the ratios between these stress-group locations will be indicated, again with respect to the different status of the latus.

Durational measurements of Type 1 disyllabics — Table 1.

	Latus has a full vowel					Latus has a reduced vowel or no vowel				
Groups	Vowel Centre $\bar{x}$	Vowel Centre SD	Consonant Centre $\bar{x}$	Consonant Centre SD	Ratios (V/C)	Vowel Centre $\bar{x}$	Vowel Centre SD	Consonant Centre $\bar{x}$	Consonant Centre SD	Ratios (V/C)
1A	192.95	20.85	227.5	32.45	0.85	233.03	28.13	266.82	26.98	0.87
2D	217.50	29.04	173.33	31.26	1.30	251.38	30.36	179.16	29.86	1.43
3B	164.37	19.04	225.00	26.69	0.73	172.50	12.37	262.50	10.50	0.65
4A	187.21	20.09	223.46	21.31	0.84	226.50	18.82	234.00	12.32	0.96
5A	234.00	22.74	246.00	22.74	0.96	221.25	26.51	303.75	10.40	0.72
6D	215.62	16.88	135.00	17.47	1.61	261.42	11.80	167.14	16.03	1.57
7B	214.50	11.37	235.50	14.62	0.91	243.75	15.90	264.31	13.63	0.92
8A	232.50	31.43	240.00	34.24	0.96	217.50	11.45	217.50	11.05	1.00
9E	233.18	23.98	186.13	12.91	1.25	255.00	10.55	207.50	10.60	1.22
10B	253.27	23.71	253.27	13.40	0.65	178.12	31.06	240.62	30.39	0.77

Durational measurements of Type 2 disyllabics — Table 2.

	Latus has a full vowel					Latus has a reduced vowel or no vowel				
Groups	Vowel Centre $\bar{x}$	Vowel Centre SD	Consonant Centre $\bar{x}$	Consonant Centre SD	Ratios (V/C)	Vowel Centre $\bar{x}$	Vowel Centre SD	Consonant Centre $\bar{x}$	Consonant Centre SD	Ratios (V/C)
1DA	170.00	15.23	365.25	34.05	0.55	211.87	21.54	361.12	31.15	0.59
2FA	156.00	33.96	337.50	33.54	0.47	150.00	11.45	345.00	8.50	0.41
3DA	136.50	22.87	312.00	10.40	0.43	150.00	10.03	322.50	11.34	0.46
4BA	120.00	10.60	367.50	10.65	0.32	123.75	10.50	402.25	8.75	0.30
5DE	132.95	15.40	322.50	33.11	0.41	163.75	13.03	365.25	26.13	0.44
6FA	150.00	15.90	382.50	12.50	0.39	153.75	12.23	401.25	10.43	0.38
7FE	142.50	10.50	302.50	24.60	0.47	162.50	12.50	347.50	10.11	0.46
8FE	165.00	12.99	292.50	38.48	0.56	150.00	8.75	300.00	24.12	0.50
9BC	142.50	11.23	322.50	10.50	0.45	130.00	10.03	395.00	9.35	0.32
10EB	165.00	25.98	317.50	33.00	0.52	172.50	8.33	322.50	21.75	0.53

Durational measurements of Type 3 disyllabics — Table 3.

	Latus has a full vowel					Latus has a reduced vowel or no vowel				
Groups	Vowel Centre $\bar{x}$	Vowel Centre SD	Consonant Centre $\bar{x}$	Consonant Centre SD	Ratios (V/C)	Vowel Centre $\bar{x}$	Vowel Centre SD	Consonant Centre $\bar{x}$	Consonant Centre SD	Ratios (V/C)
1b	292.81	36.54	68.75	17.05	4.33	337.50	12.50	75.00	10.32	4.50
2d	255.00	33.20	71.87	23.02	3.63	343.63	37.76	97.50	8.65	3.67
3e	337.50	29.45	102.50	21.41	3.29	393.75	15.05	127.00	20.30	3.10
4d	267.81	16.96	66.50	8.66	4.25	297.50	8.66	102.50	21.31	3.14
5e	284.25	32.29	77.50	18.60	3.70	340.50	6.70	87.00	19.02	3.98

Durational measurements of Type 4 disyllabics Table 4.

Groups	Latus has a full vowel					Latus has a reduced vowel or no vowel				
	Vowel Centre $\bar{x}$	SD	Consonant Centre $\bar{x}$	SD	Ratios (V/C)	Vowel Centre $\bar{x}$	SD	Consonant Centre $\bar{x}$	SD	Ratios (V/C)
1c	222.14	10.95	181.07	13.20	1.22	222.50	11.04	187.50	34.41	1.18
2b	248.25	28.39	177.45	21.88	1.41	262.50	15.44	195.00	13.33	1.35
3a	242.50	23.39	185.00	29.08	1.30	242.50	10.39	220.00	14.55	1.10
4c	210.00	27.31	185.00	34.45	1.13	228.75	12.21	191.25	23.13	1.19
5a	221.25	18.23	210.00	16.71	1.05	222.50	16.34	210.00	13.45	1.06

Durational measurements of Type 5 disyllabics Table 5.

Groups	Latus has a full vowel					Latus has a reduced vowel or no vowel				
	Vowel Centre $\bar{x}$	SD	Consonant Centre $\bar{x}$	SD	Ratios (V/C)	Vowel Centre $\bar{x}$	SD	Consonant Centre $\bar{x}$	SD	Ratios (V/C)
1fb	210.00	31.15	185.00	16.88	1.14	227.50	24.61	205.00	25.28	1.11
2da	187.50	8.61	187.50	12.11	1.00	187.50	10.12	187.50	10.41	1.00
3db	219.58	33.16	154.16	20.28	1.42	236.25	8.62	183.75	8.42	1.28
4da	210.00	12.34	157.50	23.42	1.33	242.50	9.51	175.00	12.28	1.38
5fa	277.50	30.13	142.50	31.04	1.94	285.00	12.64	157.50	23.16	1.80
6fa	235.00	23.14	187.50	30.52	1.25	253.75	22.05	227.00	27.51	1.12
7db	247.50	33.40	184.68	9.43	1.35	260.00	17.51	205.00	14.13	1.29
8fa	250.00	11.32	165.00	25.16	1.53	277.50	10.45	184.68	16.09	1.50

The mean durational difference in the vowel centre and the consonant centre in the five structural types, as well as the ratio differences with regard to the durational status of the latus, may be summarized as follows:

Mean durational differences and differences in V/C ratios in the five structural types Table 6.

Structural Types	Vowel Centre ($\bar{x}$ dif.)	Consonant Centre ($\bar{x}$ dif.)	Ratios (V/C dif.)
Type 1	34.22	28.56	0.10
Type 2	30.88	36.26	0.71
Type 3	55.09	28.70	0.33
Type 4	31.69	26.25	0.17
Type 5	34.35	29.33	0.14

The results of the measurements of the 631 Suonikylä disyllabics as summarized in the Tables raise certain concerns with regard to the importance of the quantity of the latic vowel. It seems evident that the characteristic V/C ratios that signal differences between the five structural types are not even minimally affected by the reduced quantity of the latus. As far as the attestable increase in absolute duration is concerned, this results in only incomplete compensation for

the loss of the latic duration: the sums of the durational increases in the vowel centre and the consonant centre are, in general, less than the duration of the latic vowel. It also has to be said that 23% of the test-words show little durational increase: what there is appears to be less significant than the durational increase of disyllabics summarized in Table 6. Examination of additional data (4) will involve further consideration of the relevancy of the quantity of the latus to the linguistically important quantity distribution in Skolt Sámi disyllabics. The implications of this present investigation suggest that the quantitative status of the latus may prove less significant than hitherto assumed.

Appendix

Within the five structural types, only those groups are considered in this report where a minimum of 12 disyllabics were measured. The ordering of the groups appearing in the Tables indicates the number of occurrences, commencing with the highest value.

Capital letters stand for long consonants. In the consonant clusters no indication of the durational relationship between the members of the cluster is given here. (a=stops, b=fricatives, c=affricates, d=liquids, e=nasals and f=glides).

References

1. BERGSLAND, K. : Røros-lappisk grammatikk. Instituttet for sammenlignende Kulturforskning, Oslo. 1946.
2. ITKONEN, E. : Struktur und Entwicklung der ostlappischen Quantitätssysteme. Mémoires de la Société Finno-Ougrienne 88. 1946.
3. MAGGA, T. : Duration in the Quantity of Disyllabics in the Guovdageaidnu Dialect of Lappish. Acta Universitatis Ouluensis, Series B. Oulu. 1984.
4. McROBBIE, Z. : The Duration of Disyllabics in Skolt Sámi. (in preparation).
5. SAMMALLAHTI, P. : Norjansaamen itä-Enontekiön murteen äänneoppi. Mémoires de la Société Finno-Ougrienne 160. 1977.

PARALINGUISTIC SPEECH SIGNAL TRANSFORMATIONS

Hartmut TRAUNMÜLLER and Peter BRANDERUD
Institutionen för lingvistik
Stockholms universitet

The physical properties of speech sounds are known to vary as a function of paralinguistic factors such as the speaker's age, sex, vocal effort and emotional involvement. This variation concerns also F1 and F2 in vowels. Given constant paralinguistic circumstances, however, different vowels are distinguished almost exclusively by these two formant frequencies. Considering their paralinguistic variation, it is understood that our perception of vowel quality cannot be based on these formant frequencies as such. Nevertheless, it is obvious that ordinary speech signals contain invariant correlates to the phonetic quality of vowels. It has been suggested that a process of normalization of formant frequencies guided by other vowels produced by the same speaker under similar conditions might be in effect. Although it has been shown that perceived vowel quality is affected by a preceding context in this sense (4), this does not provide an exhaustive explanation of the phenomenon. As listeners we are able to judge the phonetic quality even of a single isolated vowel, no matter by whom it has been produced, given only that we can hear the signal clearly. Therefore it must be presumed that the speech signal contains properties informative of phonetic segmental quality, free from (invariant with respect to) paralinguistic variation.

The problem of perceptual invariance despite paralinguistic variation in vowels has been the subject of a recent investigation (8). Acoustic data on vowels produced by speakers of different age and sex and in different modes of speech, including shouting and whispering, had been analysed and perceptual experiments with synthetic vowels had been performed. On this basis a description of the consequences of paralinguistic variation on F0 and the formant frequencies of vowels could be given. This also made it possible to derive cues to perceived vowel quality which are free from paralinguistic variation.

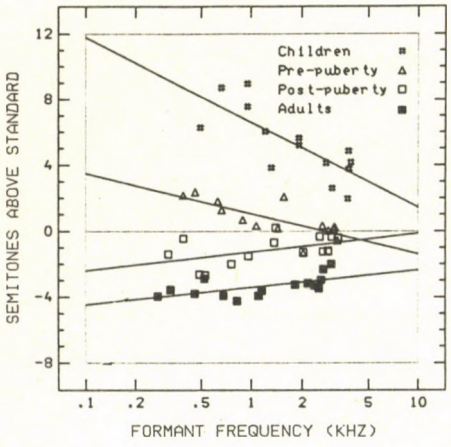


Figure 1: *Logarithmically scaled formant frequencies (F1 to F3) of the same five vowels of Japanese produced by kindergarten children, boys before maturation of voice (age 12-14 years), boys after maturation of voice (age 12-14 years), and adult men (group mean data): Deviation from a standard value (vertically) plotted against actual value (horizontally). Regression line shown for each group of speakers. Frequency data from H. FUJISAKI et al. (3).*

The relations between phonetically identical vowels produced under different paralinguistic conditions could be described by simple transformation rules. All these relations could be interpreted as linear transformations of the characteristic frequencies (F0 and F1 to F3) on a logarithmic scale (e. g. semitones) as well as on a tonotopic scale (Bark). The following exposition serves mainly to illustrate the nature of the transformation rules.

Figure 1 illustrates the ontogenetic development of vowel formant frequencies in speakers of male sex. It shows the relations between F1, F2, and F3 of the same vowels produced by kindergarten children (age 4 - 5 years), boys just before and just after maturation of voice (age 12 - 14 years for both groups) and adult men. Formant frequency is scaled logarithmically. Linear regression lines have been fit to the data from each speaker group. In this

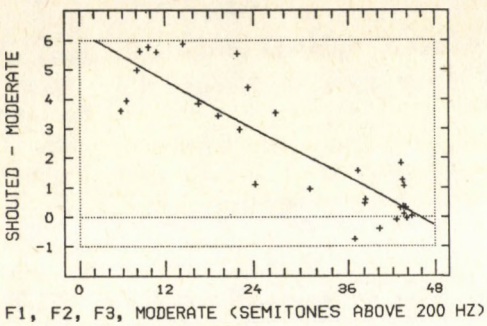


Figure 2: *Logarithmically scaled formant frequencies (F1 to F3) of the same ten Swedish vowels shouted and produced with moderate vocal effort (mean data from three speakers): Differences (vertically) plotted against values from moderate version (horizontally). Regression line (r = -0.88) also shown. Data due to R. SCHULMAN (6).*

process the formants F1 to F3 have not been discriminated. Nevertheless, the regression lines can be seen to fit the data quite well, and the remaining deviations between data points and regression lines which, not surprisingly, are largest for kindergarten children, would not be substantially reduced if we would treat the formants separately.

Figure 2 shows the ralationship between the formant frequencies in spoken and in shouted vowels produced by men. Even in this case, a linear regression line fitted to the formant frequencies without discrimination fits the data quite well. F1 is substantially increased at increased vocal effort. F2 is also affected. Formant frequency and formant frequency increase are negatively correlated. F3 remains largely unaffected. At a given degree of vocal effort, these formant frequency changes result mainly in an increased perceived vowel openness. When vocal effort is increased, this effect is offset by the increased F0 (7).

Linear relationships between logarithmically scaled formant frequencies correspond to linear regression lines in Figures 1 and 2. They are described by power functions

$$F' = k*F^p, \qquad (1)$$

in which F and F' are the frequencies (in Hz) of the same formants in the two versions, k is a measure of (vertical) displacement and p is a measure of difference in slope. The relation between different regression lines can also be described by specifying for two given frequencies, say 0.3 and 3.0 kHz, the factor by which formant frequencies change. This methode of specification has been chosen in our computer program for simulating paralinguistic variations because these figures are more immediately telling to the phonetician, as compared with the constants in equation (1). The factors of frequency change for F0 and for formants at 0.3 and at 3.0 kHz, corresponding to the regression lines in Figures 1 and 2 are listed in Table 1, in which a female - male comparison is also included.

Table 1: *Scale factors for F0 and for undiscriminated formants at 0.3 and at 3.0 kHz. Between speaker comparison based on Japanese data (FUJISAKI et al., 3), and on data from six European languages (FANT, 2). Within speaker comparisons based on Swedish data (SCHULMAN, 6).*

Transformation	k_{F0}	$k_{0.3}$	$k_{3.0}$
men - boys, mature, 12-14 years	1.29	1.14	1.13
men - boys, immature, 12-14 years	1.93	1.26	1.16
men - children, 4-5 years	2.34	1.99	1.42
men - women (Fant's data)	*	1.08	1.19
normal - shouted (men)	2.12	1.36	0.99

While it entails no advantage to discriminate between F1 and F2 in the cases shown in Figures 1 and 2, a separate treatment of these formants leads to a clearly improved description of the differences between the typical male and female realisations of the same vowel phonemes, as can be seen in Figures 3 and 4. In Figure 4, the observed average for-

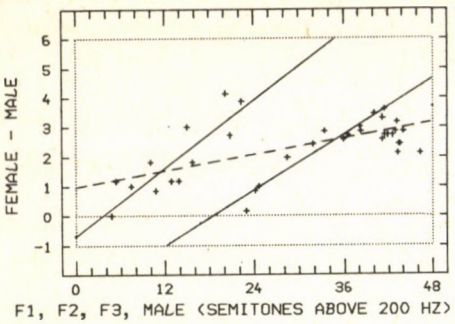


Figure 3: *Female-male differences (vertically) plotted against male values of F1 to F3 in vowels (horizontaly) on the basis of log(F). Data represent mean formant frequencies from a sample of six European languages (2). The twelve vowels chosen are those represented in at least three of these languages. Separate regression lines fitted to the data for each formant (whole-drawn lines). Overall regression line fitted to the three formants also shown (dashed).*

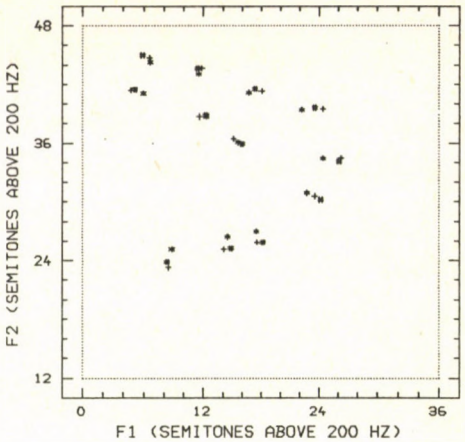


Figure 4: *F1 vs F2 diagram of female vowels. Same data as in Figure 3. Observed values (+) and values predicted from male data without discrimination (*) and with formant specific treatment (#).*

mant frequencies in vowels produced by women are compared with those predicted from male data with and without formant-specific treatment. In these Figures, women's vowels can be seen to be clearly more peripheral than men's. They are also more peripheral than those of the other three classes of speakers included in the investigation by FUJISAKI et al. (3) from which the data for Figure 1 are taken (8). The increased peripheralness, or "segmental explicity" of female vowels is akin to that of stressed vowels as compared with unstressed ones (5).

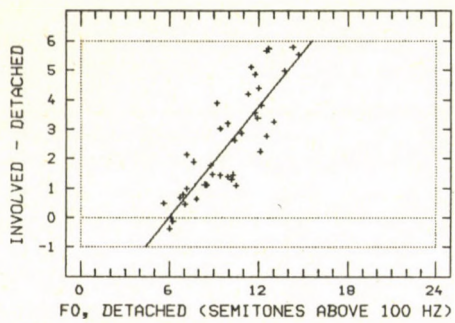


Figure 5: *Local maxima and minima in the F0-contour of four utterances produced with a detached and an involved attitude by a female speaker of Swedish. Mean values from six repetitions. F0-excess in involved version plotted against F0-values in detached version. Regression line also shown (r = 0.86). Data from G. BRUCE (1).*

There is another paralinguistic variable which we can conveniently refer to as "prosodic explicity". This variable is a corollary in several types of attitudinally and emotionally conditioned variations, reflecting the degree of personal involvement of the speaker. Increased prosodic explicity results in increased F0-excursions towards higher frequencies.

The relationship between the F0-contours of linguistically identical utterances produced with different degrees of prosodic explicity can also be described as a linear transformation on a logarithmic scale of frequency. This is shown in Figure 5, in which the local maxima and minima of F0 in some linguistically identical utterances, produced with either a low or a high degree of (faked) involvement have been plotted against each other. (The choice of scale is not crucial in this case. Linear regression lines fit the data equally well if

they are scaled in Hz or in Bark.) The speaker specific base-line value of F0 can be seen not to change. It might, however, change if vocal effect were also varied.

Now, we would like to see that the transformation rules found to hold for vowels would apply not only to vowels but to speech signals in general. In order to test this hypothesis, which constitutes the main purpose of the present study, speech signals of various kinds have been analysed with an automatic LPC-procedure in order to be resynthesized with the descriptive parameters (F0 and the formant frequencies) transformed according to the transformation rules. In doing so, the Q-values of the formants have always been conserved. In order to be able to perform such manipulations, a new program has been added to our ILS-package. This program allows to simulate the overall transformations described above. Variations in "segmental explicity" remain outside its capabilities because our LPC-analysis does not result in a reliable identification of formant number. Formant specific manipulations can, however, be performed if a frequency range can be specified within which they always apply.

These synthetic transformations should, then, bring about a change in the paralinguistic quality of the speech signal while they should leave its phonetic quality unaffected. The results obtained so far indicate this to be so as long as the transforms stay within the range of variation that can occur naturally. Thus, it is not possible to lower the formant frequencies below the values they have in the speech of adult men without affecting phonetic quality. An excessive lowering of F2 and the higher formants will add an overall labialized quality to the speech. In the frequency range above 2 kHz it is not either possible to increase the formant frequencies above the values they have in our kindergarten children, otherwise front rounded vowels will loose their roundedness and some consonants will aquire a palatalised quality. F1 and, to a lesser extent, a low F2 can apparently be increased to higher values than in the normal speech of kindergarten children. A minor problem that has been observed with transformed consonants consists in improper intensity levels, noticeable mainly in the "voice bar" of voiced stops. Within the mentioned limitations, the transformations described above apply to consonants as well as to vowels. The remaining deficiencies of the transformed speech have mainly to do with those properties of the speech signal which in a traditional framework are ascribed to the voice source.

Acknowledgment: This research has been supported by a grant from HSFR, the Swedish Council for Research in the Humanities and Social Sciences.

References

1. BRUCE, G.: Developing the Swedish intonation model. Working Papers 22, Lund university, Department of linguistics, 1982, 51-116.
2. FANT, G.: Non-uniform vowel normalization. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., No. 2/3 pp. 1-19 (1975).
3. FUJISAKI, H.; YOSHIMUNE, N.; NAKAMURA, N.: Formant frequencies of sustained vowels in Japanese obtained by analysis-by-synthesis of spectral envelopes. (Unpulished data, Univ. of Tokio, 1970).
4. LADEFOGED, P.; BROADBENT, D. E.: Information conveyed by vowels. J. Acoust. Soc. Am. 29: 98-104 (1957).
5. RIETVELD, A. C. M.; KOOPMANS - VAN BEINUM, F. J.: Vowel reduction and stress. Speech Comm. 6: 217-229 (1987).
6. SCHULMAN, R.: (Unpublished data, Stockholm Univ., 1985)
7. TRAUNMÜLLER, H.: The role of the fundamental and the higher formants in the perception of speaker size, vocal effort, and vowel openness. PERILUS, No. 4, (Inst. Linguist., Univ. Stockholm), pp. 92-102 (1985).
8. TRAUNMÜLLER, H.: Paralinguistic variation and invariance in the characteristic frequencies of vowels. Phonetica 45: 1-29 (1988).

WORTPROSODIE IM AUSSAGESATZ DER LITAUISCHEN SPRACHE

Valerija VAITKEVIČIŪTĖ
Staatliches Konservatorium, Vilnius, Litauische SSR

Material

Die Akzente der litauischen Hochsprache werden in 132 Rahmen-Aussagesätzen untersucht, in die die Quasi-Homonyme (gí:vi) - (gĩ:vi), (sú:ri) - (sũ:ri), (a:ustre) - (áu:stre) eingesetzt werden. Abhängig von der Sinnbetonung und der Stellung des betreffenden Wortes in der Aussage entstehen 6 Satztypen - I, II, III, IV, V, VI, von denen je ein Beispiel hier angeführt wird: I. Septintas linksnis yra (gí:vi). "Der siebte Kasus ist (gí:vi)". II. Septintas linksnis yra (gí:vi). III. Septintas, linksnis yra (gí:vi). IV. Septintas linksnis yra (gí:vi). V. (gi:vi) yra spetintas linksnis. '(gí:vi) ist der siebte Kasus". VI. (gí:vi) yra septintas linksnis.

Informanten, Methodik, Arbeitszweck

Die Sätze wurden von Informanten, Vertretern verschiedener Mundarten (3), in Hochlitauisch auf das Tonband gesprochen. Auf Grund der Tonaufnahmen wurden die Oszillogramme der Sätze angefertigt. Die Akzente werden nach derselben Methodik wie auch bei Einzelwörtern erforscht (3). Zwecks der Audioforschung wurden die betreffenden Wörter aus dem Satz geschnitten und paarweise zusammengeklebt. Haben sich die Hörer ein Paar angehört, mußten sie feststellen, ob sich die Vokale in diesen Wörtern unterscheiden oder nicht, wenn ja, dann wodurch. Dabei wird das Ziel verfolgt, nachzuweisen, wie die vom Satztyp abhängig Wechselwirkung zwischen Satzintonation und Akzent sehr verschieden ist, was von der dialektalen Zugehörigkeit des hochlitauisch sprechenden Informanten im wesentlichen bedingt wird (3).

Die Wahrnehmung der Akzente in Sätzen

Die Akzente der zusammengesetzten Diphtohonge in den betreffenden Wörtern werden von allen Hörern sehr leicht wahrgenommen und richtig bezeichnet: ′,~. Dafür ist der in allen Satztypen erhalten gebliebene wesentliche Unterschied in der Dauer /t/ der ersten Komponenten der Diphthonge mit unterschiedlichen Akzenten entscheidend. Daraus ergeben sich ein sehr unterschiedliches Verhältnis von t der Diphthongkomponenten und eine ganz unterschiedliche Qualität der Komponenten. Andere für die Wahrnehmung der Akzente der Monophthonge wichtige Faktoren haben für die Diphthonge so gut wie keine Bedeutung. Die Wahrnehmung der Akzente der Monophthonge in den betreffenden Wörtern im Satz hängt davon ab, ob der wesentliche Unterschied bei den Hauptparametern erhalten bleibt oder nicht. Bleibt im Satz nur der wesentliche Unterschied der Dauer derselben Monophthonge mit unterschiedlichem Akzent erhalten, so werden sie von den Hörern nur wie Laute von unterschiedlicher t wahrgenommen. Damit die Hörer die Monophthongakzente wahrnehmen und sie richtig mit ′,~ bezeichnen, gelten folgende Voraussetzungen: 1) außer dem wesentlichen Unterschied in der Dauer der Monophthonge muß der wesentliche Unterschied des zweiten von der dialektalen Zugehörigkeit des Informanten abhängigen Hauptparameters erhalten sein: bei D1, D4 - der Unterschied in der Intensität (J); bei D2 - im Hauptton (Ht); bei D5 - der Unterschied in der Summarenergie (SE) der nachtonigen Monophthonge sowie der Intervall-Unterschied des Haupttons des betonten und

nachtonigen Vokals; 2) das Sinken des Haupttons des betreffenden Wortes mu3 die informativen Teile des Monophthongs nicht erfassen.

Die Modifikation der Akzente in Sätzen

Wie die Wahrnehmung der Akente ergibt, ist die Modifikation der Monophthongakzente wegen Satzintonation dreifach: 1) die volle Nivellierung der Akzente, wenn der t-Unterschied der Monophthonge sich verschlechtert oder völlig eliminiert ist, während der Ht der informativen Teile des Monophthongs absinkt (D1, D4 IV; D4 I/; 2) die Verstärkung der Gegenüberstellung der Akzente, wobei der Unterschied in den Hauptparametern (3) wesentlich bleibt, während der Ht entweder überhaupt nicht absinkt (D1 V) oder sein Sinken die informativen Teile des Monophthongs nicht betrifft (D2 III); 3) die Teilnivellierung der Akzente, wenn der wese liche t-Unterschied der Monophthonge erhalten bleibt, und der Ht des ganzen Monophthongs absinkt (D1 I, II, III; D2 I, II, IV; D4 III; D5 V). Anschlie3end werden zwei erstere Akzentmodifikationen näher behandelt.

1) Die volle Nivellierung der Akzente (D1, D4 IV; D4 I)

D1 IV. In den Sätzen des Informanten D1 IV wird für die Gegenüberstellung der Akzente der viel schwerer zu vernehmende Unterschied bei den dynamischen Mitteln, d.h. in der Intensität und Vergleichsenergie (VE) angewendet. Im Vergleich zu den Einzelwörtern bewahren den dadurch ausgedruckten Unterschied die Monophthonge und Diphthonge in allen Paaren auf, Monophthonge und nur die ersten Komponenten der Diphthonge sowie die III. informativen Teile der Vokale, während die I. Teile sich nur bei den Monophthongen unterscheiden: der I-Unterschied bleibt in allen Fällen wesentlich (Abb.1), vgl. Fig. 1(3). Der t-Unterschied der Monophthonge wird modifiziert: im 1. Paar ist er länger im Akut (6%), im 2. - im Zirkumflex (4%). Der Ht-Unterschied bleibt nur bei den I. Teilen des 1. Paares erhalten. Die Akzente der Monophthonge werden völlig nivelliert wegen der Modifikation des t-Unterschieds und des Sinkens des Ht, das deren I. informativen Teil betrifft.

D4 IV. Der t-Unterschied der betreffenden Wörter bleibt nur bei den ersten Komponenten der Diphthonge erhalten (62%), Akute Monophthonge sind länger als Zirkumflexe - 24%, 29 %. Den wesentlichen J-Unterschied (19%) bewahren die gesamten Monophthonge nur im 2. Paar auf, und bei den II. Teilen ist der Unterschied unwesentlich (9%, 6%) - im 1. bzw. 2. Paar (Abb. 2), vgl. Fig. 2(3). Den VE--Unterschied weisen nur die II.Teile im 1. Paar auf, der wesentliche Ht-Unterschied ist für die III. Teile in allen Paaren und für die II. Teile im 2. bzw. 3. Paar kennzeichnend. Obwohl der Anteil des Ht in der supersegmentalen Ebene grö3er ist als in der segmentalen, werden jedoch die Akzente wegen des Ht-Sinkens, das die II. und III.-informativen Teile erfa3t, und wegen des t-Unterschieds völlig nivelliert.

D4 I. Der t-Unterschied ist im 1. Paar unwesentlich (4%), im 2.- wesentlich (18%); im 1. bzw. 2. Paar bleibt der der wesentliche I-Unterschied bei den I. Teilen bestehen (23%, 29%), im 1. Paar ist er bei den gesamten Monophthongen unwesentlich (2%), im 2. Paar - wesentlich (16%); der wesentliche I-Unterschied der II. Teile (19%) ist nur im 1. Paar erhalten (Abb. 3), vgl. Fig. 2(3). Die I. Teile unterscheiden sich durch

ihre VE in allen Paaren, die II. Teile - im 1. bzw. 2. Paar. Der Ht-Unterschied der Monophthonge bleibt nicht erhalten, dazu betrifft sein Sinken deren informative, d.h. die II. und III. Teile.

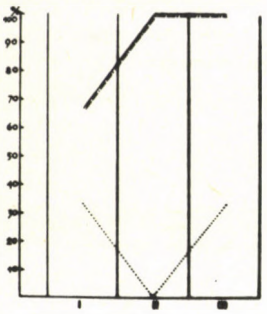


Abbildung 1. D1 IV.

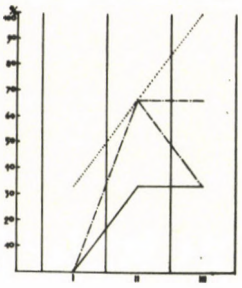


Abbildung 2. D4 IV.

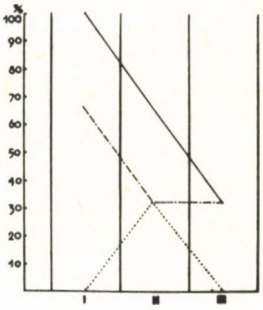


Abbildung 3.D4 I.

Gegenüberstellung der Akzente in Aussagesätzen: .... der wesentliche Ht-Unterschied; -.-.- der unwesentliche ϑ-Unterschied; --.--.-- der wesentliche I-Unterschied; ——— der VE-Unterschied.

2) Die Verstärkung der Gegenüberstellung der Akzente (D1 V; D2 III)

D1 V. Der wesentliche t-Unterschied bleibt in allen Paaren erhalten. Den wesentlichen ϑ-Unterschied bzw. den VE-Unterschied weisen nur die gesamten Monophthonge im 1. und 2. Paar, d.h. deren I. bzw. III. Teile, auf. Der wesentliche ϑ-Unterschied dieser informativen Teile ist in der supersegmentalen Ebene grö3er (47%-73%) als in der segmentalen (9%-40%). Abb. 4, vgl. Fig 1(3). Der Ht der I. Teile ist jedoch in allen Paaren höher als der der III. Teile, der Ht der III. Teile ist höher als der Ht der nachtonigen Monophthonge, der Ht der II. Teile ist nicht immer tiefer als der der I. Teile. Daraus ergibt sich, da3 das gesetzmä3ige Sinken des Ht ist für das ganze Wort und nicht für den Monophthong kennzeichnend. Alle Höhrer bezeichneten die Akzente der Monophthonge mit ˊ,~.

D2 III. Alle Paare weisen den wesentlichen t-Unterschied auf. Der wesentliche ϑ- bzw. VE-Unterschied ist für die gesamten Monophthonge und Diphthonge, für die Monophthonge und nur für die ersten Komponenten der Diphthonge sowie für die I. Teile der Monophthonge typisch (Abb. 5), vgl. Fig 3(3). Der Ht-Unterschied bleibt bei den gesamten Monophthongen und Diphthongen in allen Paaren, bei Monophthongen und nur den ersten Komponenten der Diphthonge und deren informativsten I. und II. Teilen erhalten, und die III. Teile bewahren ihn nur im 1. Paar auf. So ist der Unterschied aller Parameter klarer ausgedrückt als im Falle der Einzelwörter. Die Vokale mit unterschiedlichem Akzent weisen weder das gesetz-

mä3ige Sinken noch das Steigen des Ht auf. Der Ht des nachtonigen Monophthongs ist tiefer als der Ton des beliebigen Teils eines betonten Vokals. Daraus folgt, da3 das Sinken des Ht die informativen Teile des betonten Vokals nicht betrifft.

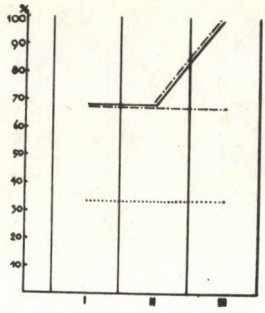


Abbildung 4. D1 V.
Erläuterungen s. zu Abb. 1,2,3.

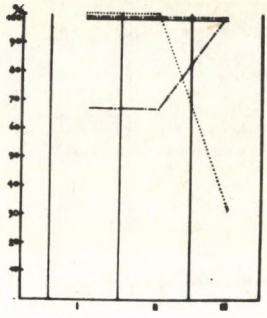


Abbildung 5. D2 III.

Schlu3folgerungen

Die vorliegende Untersuchung trägt eine Berichtigung in die Schlu3folgerungen manchen Autoren ein. Nur in den Sätzen vom Typ IV. der Informanten D1, D4 wird der t-Unterschied bei denselben Monophthongen mit unterschiedlichem Akzent völlig nivelliert. In der Abhandlung (1) wird behauptet, da3 der t-Unterschied in der Phrase immer nivelliert wird. Dieser ziemlich kategorische Schlu3 ergab sich, da man bei der Feststellung des t-Unterschieds die Stellung des betreffenden Wortes im Satz sowie die Spezifik des Aussage- bzw. Fragesatzes nicht in betracht zog. In der Abhandlung (2) wird t der zusammengesetzten Diphthonge ganz unbegründet für ein invariantes Kennzeichen von deren Akzenten gehalten.

Literatur

1. ANUSIENĖ, L.: Duration of long stressed vowels in Present-day Lithuanian utterances. Proc. XIth ICPhS 5. 1987, 99--102.
2. PAKERYS, A.: Lietuviu bendrines kalbos fonetika. Vilnius, 1986.
3. VAITKEVIČIŪTĖ, V.: Pitch accents in Standard Lithuanian. Proc. XIth ICPhS 5. 1987, 103--106.

GEDANKEN
ÜBER DIE GESCHICHTLICHE VERÄNDERUNG DES SPRECHTEMPOS

András O. VÉRTES
Institut für Sprachwissenschaft
der Ungarischen Akademia der Wissenschaften
Budapest, Ungarn

1. Hat sich das Sprechtempo gegenüber dem Zustand vor Jahrtausenden geändert - und wenn ja, wie läßt sich dies beweisen? Hat sich das Sprechtempo in manchen Sprachen im Laufe der letzten Generationen beschleunigt? Läßt sich diese jüngste Beschleunigung in Kulturbereichen jenseits der Sprache nachweisen? Hängt die Veränderung des Sprechtempos mit der Änderung der Stimmlage der Rede zusammen? - ein Teil dieser Fragen soll nun in erster Linie aus der Sicht der ungarischen Sprache behandelt werden.

2. In manchen Sprachen hat sich das Sprechtempo gegenüber dem Zustand vor Jahrtausenden geändert. Deutlich wird dies, wenn sich das Sprechtempo von Einzelsprachen unterscheidet, die von der gleichen Grundsprache abstammen. So läßt sich im Falle der zur indoeuropäischen Sprachfamilie gehörenden Sprachen schnelleren (Französisch und Spanisch) und langsameren (Deutsch) Typs feststellen, daß sich das Sprechtempo zumindest einer seit der Zeit der Grundsprache geändert hat. (Ähnlich urteilt Kubínyi in seinem vorzüglichen Artikel: 1958).

János Hunfalvy schrieb vor mehr als hundert Jahren, daß die Finnen langsam sprachen, ebenso wie die Ungarn, die Esten dagegen sehr schnell (in einem Brief: Reval, 9. VII. 1869; auf meinen Wunsch hat Viljo Tervonen die betreffenden Briefabschnitte kopiert). Auch nach Beobachtung von Professor Ariste wird im Estnischen schneller als im Finnischen gesprochen (mündliche Mitteilung vom 21. IX. 1960).

Es scheint so, als werde auch in einer anderen finnisch-ugrischen Sprache, dem Mordwinischen, schneller gesprochen als im Ungarischen, zumindest; ist dies die Beobachtung des mordwinischen Sprachwissenschaftlers Fedor Markow (mündliche Mitteilung vom 7. VII. 1961).

Also hat sich das Sprechtempo auch seit der finnisch-ugrischen Grundsprache geändert.

(Hier ließe sich einwenden, im Prinzip könnte man bereits in der Grundsprache dialektale Tempounterschiede voraussetzen.)

3. Nach Wundt hatte sich auch schon bis zu seiner Zeit das (deutsche) Sprechtempo beschleunigt, "für das uns unmittelbare Zeugnisse kaum zu Gebote stehen" (1904,[2] I. Teil, 488). Ein "charakteristisches Zeugnis" fand er in der Tempobeschleunigung der Musik. "Bekanntlich hören wir selbst Beethovens Symphonien heute in der Regel in einem schnelleren Tempo vorgetragen, als in dem sie ursprünglich komponiert waren; und noch größer ist dieser Unterschied bei Meistern wie Haydn oder Mozart, Händel oder Bach..." (ebd.; vgl. noch S. 489). Wundt zufolge lassen auch die grammatischen Formen auf "ein langsameres, sozusagen majestätisch einherschreitendes Tempo der (gothischen und althochdeutschen) Rede schließen" (ebd. 489).

Im folgenden wird versucht, einerseits unmittelbarere Beweise für die Tempobeschleunigung zu geben und andererseits zu belegen, daß das Sprechtempo (zumindest in einigen Sprachen) auch seit der Jahrhundertwende eine Beschleunigung erfahren hat.

Schon der Psychologe Willy Hellpach sprach von der Erscheinung der

schnellen Gro3stadtsprache, der Stenolalie (1952, 99).

Weltweit leben immer mehr Menschen in Städten, folglich verbreitet sich die Stenolalie ständig. Auch Radio und Fernsehen gewöhnt die langsamer sprechenden Provinzbewohner immer mehr an das Sprechtempo der Gro3stadt.

Wir vermuten nicht nur, sondern besitzen auch - in gewissem Ma3e direkte - Beweise, da3 sich das Sprechtempo im Laufe der letzten Jahrzehnte gesteigert hat: die Notizen der Stenographen. (Solche verwendete bereits Marbe zu Anfang des Jahrhunderts.)

Zu Beginn der zwanziger Jahre schrieb Gyula Nosz, das allgemeine Redetempo im ungarischen Parlament sei auf 240-250 Silben pro Minute angestiegen (1924, 110). Während zahlreiche Abgeordneten in einer Parlamentsdebatte von 1881 die Geschwindigkeit von 180 Silben nicht erreichten, fand sich in der ersten Hälfte der zwanziger Jahre kein einziger Redner mit einem solchen Tempo (ebd.; vgl. Vikár 1889, 2-3; vgl. noch Nosz 1925). Dieser Wandel scheint noch auffälliger zu sein, wenn wir die Bemerkung aus den ungarischen stenographischen Blättern 1869 akzeptieren, wonach "der mit mä3igem Tempo sprechende Redner pro Minute 120-140 Silben ausspricht" (Markó 1869, 3). Nach meinen Beobachtungen und Messungen empfinden wir einen Vortrag im Tempo 140-150 Silben bereits als schleppend. Wenn auch "die übereinstimmende Feststellung alter erfahrener Stenographen und besonders der Parlamentsstenographen" übertrieben ist, wonach sich innerhalb eines Vierteljahrhunderts bis zum Beginn der zwanziger Jahre das Tempo öffentlicher Reden fast auf das Eineinhalbfache erhöhte (Blasovszky 1922, 19), ist eine erhebliche Beschleunigung nicht zu bezweifeln.

Auch die deutsche Parlamentsrede hat sich beschleunigt. Das Dresdener Stenographische Landesamt antwortete am 14. August 1958 brieflich auf meine Frage: "Die Redegeschwindigkeit in den Parlamenten hat sich in den letzten Jahrzehnten immer mehr erhöht. Man kann sagen, da3 eine ständige Steigerung der Redegeschwindigkeit zu verzeichnen ist. Auch in der Zeit von 1945 bis zur Gegenwart stieg die Redegeschwindigkeit in unseren Parlamenten.

Die durchschnittlichen Redegeschwindigkeiten liegen heute bei 240/260 Silben je Minute. Solche verhältnismä3ig hohen Geschwindigkeiten gibt es nicht immer...

Um die Jahrhundertwende etwa mag die Redegeschwindigkeit bei etwa 200/220 Silben je Minute gelegen haben. Höhere Geschwindigkeiten sind bestimmt eine Ausnahme gewesen."

Das Stenographische Amt der Volkskammer der DDR hat ebenfalls auf meine Fragen geantwortet (Berlin, 27. VIII. 1958): Was die "Frage nach der Redegeschwindigkeit im Parlament betrifft, so schätzen wir, dass die Redegeschwindigkeit zwischen etwa 240/320 Silben in der Minute schwankt...

... Selbstverständlich wechselt das Redetempo mancher Redner stark. So kommen gelegentlich auch Geschwindigkeitsspitzen nach unserer Schätzung von 340-360 Silben für kurze Perioden vor. Ihre Anfrage wird uns Anregung sein, einmal im Parlament genauere Messungen anzustellen. Sollten die Ergebnisse von den vorgenannten Zahlen abweichen, werden wir sie Ihnen übermitteln. Im allgemeinen vertreten Parlamentsstenographen die Meinung, da3 die Redegeschwindigkeit sowohl in als auch au3erhalb der Parlamente in den letzten Jahrzehnten zugenommen hat." Die hier mitgeteilten Daten beruhen sich aber nicht auf "fundierten Angaben".

Die Redegeschwindigkeit im Theater ist nicht nur ein Teil, sondern auch ein Spiegel des allgemeinen Sprechtempos. Auch deshalb beansprucht die Feststellung des Dramahistorikers und Dramaturgen Sándor Galamb (1886-1972) Aufmerksamkeit, wonach die Beschleunigung des Sprechens auf der ungarischen Bühne besonders in den 1910er und 1920er Jahren spürbar wurde; ebenfalls von ihm erfuhr ich, da3 das Provinzschauspiel immer "bequemer" ablief

(mündliche Mitteilung 1958 oder 1959), ich vermute, daß dieser Tempounterschied auch mit der Erscheinung der Stenolalie in der Großstadt im Zusammenhang steht. (Zu den Gründen für die historische Veränderung -- der Beschleunigung oder der Verlangsamung -- des Sprechtempos gehören unter anderem auch die Lautwandel: im Altungarischen hat sich z.B. die Aussprechzeit der kurzen Vokale in vielen Fällen verlängert, auch wenn der kurze Vokal eigentlich nicht zum langen Vokal wurde, so im Falle des Wandels [i] > [e] und [u] > [o]; nicht unberücksichtigt darf unter den Gründen des Sprechtempos das persönliche psychophysische Tempo der einzelnen Mitglieder einer Sprachgemeinschaft bleiben.)

4. Das Redetempo kann auch die Verständlichkeit der Rede beeinflussen. Noch 1954 wandte man sich mit dem Wunsch an das Ungarische Radio, daß die Reporter in den Dorfsendungen langsamer sprechen mögen (Magyar Nemzet 13. XI. 1954, 4). Interessant ist die Rückerinnerung einer ungarischen Psychologin an ihre Schuljahre: als sie aus der Gemeinde Gúta im Komitat Komárom in die Stadt Szolnok umzog, verstand sie manchmal gar nicht, was die Lehrer und Schüler sagten, weil deren Sprechtempo viel schneller war als in Gúta; ihre Mitschüler wiederum mußten über ihr langsames Sprechen lächeln; sie fügte hinzu, daß das Sprechtempo der Kinder in Gúta ungefähr dem ihren entsprach (mündliche Mitteilung von Edit S. Molnár etwa 1960).

Hierher gehört ebenfalls, daß unter den ungarischen Theaterhistorikern die Ansicht nicht unbekannt ist, zur Zeit des ausgezeichneten ungarischen Schauspielers Gábor Egressy (1808-1866) hätten die Schauspieler langsamer sprechen müssen, da das Publikum weniger Kultur besaß (mündliche Mitteilung von Edit M. Császár im Mai 1960).

5. Nicht zu vernachlässigen ist die Erscheinung, daß auch das Tempo des ungarischen Gesangs sich geändert hat: die heutige Dorfjugend singt schneller als die entsprechende Generation im alten Dorf: mündliche Mitteilung des Musikologen Benjamin Rajeczky (1901-) am 17. Mai 1960.

Im kulturgeschichtlichen Hintergrund der erwähnten Tempoveränderungen finden sich -- parallel mit ihnen -- auch andere Erscheinungen. Eine solche ist die Beschleunigung von Musik und Tanz auch seit der erwähnten Untersuchung von Wundt, also nach der Jahrhundertwende.

Der französischen Musikologin Gisele Brelet zufolge hat das Tempo des musikalischen Vortrags seit dem letzten Jahrhundert ständig und seit 1900 sehr schnell zugenommen ("la vitesse d'execution s'est très rapidement accrue", in einem Brief an mich vom 1. September 1959). Auch nach Ansicht des französischen Musikologen Daniel Lazarus war das Tempo des musikalischen Vortrags zu Beginn der 60er Jahre schneller als 60 Jahre zuvor (Mitteilung von Mme Denise Lazarus in einem Brief vom 15. Februar 1961, die Ansicht ihres schwerkranken Gatten wiedergebend).

1944 beklagt sich ein unbekannter Autor darüber, daß das Tempo der Tanzmusik und so auch der Tanz sehr viel schneller geworden seien (Veszprémi Hírlap 2. VII. 1944, 3). Bence Szabolcsi hält für das größte Übel beim Csárdás, daß er in seiner "frischen" Form seit 1853 zu schnell gespielt werde, zu einem wahren Cancan geworden sei (Bence Szabolcsi 1961, II, 208).

6. Gibt es einen Zusammenhang zwischen der Beschleunigung der Rede und der Veränderung ihrer Stimmlage? Nach Meinung des Musikologen Máté Pál sprechen Menschen mit Baßstimmlage langsamer. Der Musikgelehrte Pál Járdányi hält es für möglich und Benjamin Rajeczky für wahrscheinlich, daß ein schnelleres Sprechen eine höhere Stimme verlangt. (Alle drei in mündlichen Mitteilungen vom Mai 1960.)

Einer der bahnbrechenden Theoretiker der ungarischen Theaterliteratur, Gábor Egressy schrieb, mit der tiefen Stimme pflege eine langsame Rede in Verbindung zu stehen, während die hohe Stimme die Beschleunigung der Rede

mit sich bringe (1879, 128).

Da3 die tiefere Stimme im allgemeinen langsamer sei, behaupteten auch die Theaterhistoriker Edit M. Császár und Géza Staud (mündliche Mitteilung im Mai 1960.).

Damit hängt zusammen, da3 die pathetische Stimme nicht nur tiefer ist (Edit M. Császár), sondern auch langsamer (Sándor Galamb, mündliche Mitteilung um 1958).

Mit all dem wurde demnach nicht nur bewiesen, da3 sich die ungarische und die deutsche Sprechgeschwindigkeit auch seit der Jahrhundertwende beschleunigt hat, sondern auch wahrscheinlich gemacht, da3 sich die Stimmlage unseres Sprechens -- mehr oder weniger parallel mit der Beschleunigung -- erhöhte.

Literatur

1. BLASOVSZKY Miklós: A gyorsírás határa: Az Irás 13. 1922. jan.--május, 19.
2. EGRESSY Gábor: A szinészet iskolája, Budapest 1879.
3. KUBÍNYI László: Magyar nyelvtörténeti változások vélhető összefüggéseiről. Magyar Nyelv LIV. 1958, 213--232.
4. MARKÓ Sándor: Gyorsirászati próbatét. Gyorsirászati Lapok 7/5--6. sz. 1869, 3.
5. NOSZ Gyula: A nemzetgyűlés szónokai. Az Irás 15. 1924, 22-24, 107--111.
6. NOSZ Gyula: Szónokaink és a magasfoku gyorsírás. Az Irás 17. 1925. 17--20.
7. SZABOLCSI Bence: A magyar zene évszázadai. II. Budapest 1961.
8. VIKÁR Béla: A magyar szónoklat sebessége (Statisztikai adatokkal). Gyorsirászati Lapok 27/10. sz. 1888/1889, 2--3.
9. WUNDT, W.: Völkerpsychologie. I. Band. Die Sprache, I--II. Teil. Leipzig 1904^2.
10. Hellpach, W.: Mensch und Volk in der Gro3stadt. Stuttgart 1952^2.

DER KOMPLEX VON SITUATION, TEXT UND ABSICHT ALS DETERMINANT DES KLANGES DER REDE

Imre WACHA
Institut für Sprachwissenschaft
der Ungarischen Akademie der Wissenschaften

1. Die Mittel der Intonation, welche wir Mittel der Satz- und Textphonetik zu bezeichnen pflegen, erscheinen im Satz, innerhalb des Satzes in den Wörtern, Wortelementen (Silbes), so werden sie realisiert, ihren Sinn erhalten sie aber erst im Zusammenhang des Textes, ja sogar des Kommunikationsfeldes.[1]

Die Anwendung der Intonationsmittel wird nämlich im gesamten Sprechprozess durch Mittel oberhalb des Satzes, sehr oft außerhalb des Satzes bestimmt, die auch den Sinn ergeben. Faktoren, welche in der Regel auch den Text, die Textform (Gattung, Genre, Stil) bestimmen. Bei der Entstehung des Textes bestimmen sie den Prozess der Textbildung, in anderen Fällen des Textklanges (der Textertönung), in wiederum anderen Fällen den Textbildung und -ertönung.

2. Wir wissen, daß die Sprache, die langue zwei Hauptrealisierungsformen hat, die geschriebene Parole und die gesprochene Parole.[2]

2.1. In die große Kategorie der geschriebenen Parole können wir Produkte der schriftlichen Kommunikation einstufen. Hauptmerkmal der geschriebenen Parole ist, daß die Konzipierung und Festhaltung der Gedanken einen relativ längeren und meistens vertieften schöpferischen Prozess oder eine Handlungsreihe darstellen. Dabei wird vom Absender "denkend konzipiert".[3] Die so konzipierten Schriftwerke charakterisiert - sprachlich - die Vielzahl von kompakten sog.simultanen Sätzen. Diese sind förmlich oft einfache oder wohlproporzioniert redigierten zusammengesetzten Sätze, in denen es viele komplizierte, auch Neben- und Unterordnungen enthaltende, aus 7-8, manchmal 10-14 Wörtern bestehende Wortkonstruktionen gibt.[4]

2.2. Bei der gesprochenen, genauer bei der verklingenden Parole unterscheiden wir zwei Hauptvariation der Realisierung. Eine von diesen ist das lebendige Sprechen, die spontane Rede.

2.2.1. Von den Merkmalen des Textbildungsprozesses und somit seiner Klangwelt ist das wichtigste, daß der Sprecher (Sender) - mindestens beim Beginn der Interaktion - nur im Besitz seiner Sprechabsicht, seines Sprechziels, sowie einer "globalen Mitteilung" ist. Der zur Übergabe der Mitteilung notwendige sprachliche Weg und sprachliche Form wird von ihm während der Interaktion gesucht-geschöft, oft etwa gemeinsam mit dem Gesprächspartner (dem Abnehmer).

2.2.2. In einem solchen Text (Sprechwerk) gibt es viele unvollständige oder unredigierte Sätze. Diese erhalten ihren Sinn oft nur durch die gemeinsamen Vorkenntnisse der Gesprächspartner, durch die gemeinsamen vorangehenden Texte und die Sprechsituation.

2.2.3. Es gibt viele Sätze, ja sogar Satzreihen, die - hauptsächlich

grammatisch - nicht korrekt, genau konstruiert sind, die als Ergebnis des "konzipierenden Denkens" voll mit offengelassenen Konstruktionen, sich den Äußerungen anknüpfenden nachträglichen Hinzufügen, Ergänzungen voll sind. So geht die Konstruktion des Satzes (?) renkend, mit komplizierten Biegungen, Rückkoppelungen voran.

All das erweist sich im Klang so, daß die Abgeschlossenheit der Sätze außer der sprachlichen Form auch durch Pause signalisiert wird, bleibt die Mehrheit der Sätze akustisch, hinsichtlich ihrer Intonation offen. Die offengehaltene, nach vorn weisende Intonation signalisiert, daß der Sprecher noch nicht am Ende der Gedankenreihe steht und fortsetzen will — wenn auch nicht unbedingt den Satz, sondern — das Sprechen. Die Intonation verspricht die Fortsetzung der Äußerung, sagt aber weder den Charakter, die Struktur noch ihre Abgeschlossenheit voraus. Oft deutet jedoch die Intonationsform der Fortsetzung darauf, daß eigentlich eine neue Äußerungseinheit began.

2.2.4. In der spontanen Rede gibt es viele Äußerungen, deren Einheiten grammatisch nicht akustisch aber wohl abgeschlossen sind. Die Intonation vielmals endet auf dem Tiefpunkt, und an dem wird mit einem neuen akustischen Start und mit Hilfe des Sprechrhytmus irgendeine nachträgliche Ergänzung, angeknüpt: Manchmal der Schlußteil des Satzes. Die Gesamtmitteilung ist also grammatisch ein mehr oder weniger perfekt konstruirter Gesamtsatz, akustisch jedoch eine Reihe von Äußerungseinheiten von mangelhafter Struktur, deren Kohesion neben gewissen grammatischen zusammenhaltenden Elementen (gemeinsames Subjekt, Suffixe usw.) durch rhytmische Elemente der Sprache (Drauschlagen auf die Vorgeschichte) und eine eigenartige, zurückschaltende Intonation signalisieren. Sehr häufig ist der intinative Abschluß "während des Satzes", während in Intonation am Satzende unabgeschlossen ist.

In der spontan Rede gibt es viele Sätze, sogar Satzreihen, die grammatisch nur mehr oder weniger, vor allem jedoch akustisch abgeschlossen sind. Diese mit Pausen "einfallmäßig" gegliederten Äußerungen werden von akustisch neugestarteten, grammatische aber nicht vollständigen Mitteilungen mit höchstens Satz konstruktions-, Gliedsatz- oder Nebensatzwert gefolgt.

3. Da bisher gesagte ergibt folgende akustische Merkmale der spontanen Rede:

a) die Rede enthält viele offengehaltene Intonationsformen (sog. Sprungschluß; am Ende hebend), und die ganze Information wird von abwechslungsreichen Intonationsformen, Melodiegägen begleitet in Betonungs- und Tonfalleinheiten, die den Rhytmus oder die Flottheit-Abgerissenheit des Denkens widerspiegeln. Die Melodie, dern Rhytmus, die Dynamik werden von der Attitüde weiter gefärbt. Dies ist im Ungarischen deshalb von Interesse, weil — wie wir es bisher registriert hatten — die eigenartigste und häufigste Melodieform der ungarischen Sprache der sog. vornfallende. Tonfall ist.

b) Die Satzgrenzen sind manchmal volkommen unfeststellbar. Die Gesamtmitteilung, der Gesamttext stellt sich aus komplizierten Ketten von Sätzen und Satzeinheiten, aus "Sätzen" mit außerordentlich komplizierten Strukturen, die meisten successiv konstruiert sind, zusammen. Trotzdem, daß es zahlreichte, in der Tiefe auf mehreren. Eben gegliederten Gesamtsätze gibt, allgemein ist einerseits die in zwei Ebenen gegliederte Satzkonstruktion, andererseits — bei Sätzen mit mehreren Ebenen — die kettenartige Satzkonstruktion, deren Satzeinheiten nicht unter das

Niveau der Satzfunktion gelangen. Dies resultiert im Klang in einer anderen Intonation und in einem eigenartig pulsierenden Sprechrhytmus.

c) Die rhytmische Gliederung der Rede bzw. die Anwendung von Pausen kennzeichnet, daß durch die gliedernden oder Hesitationspausen die sprachlich-logische Einheit oft verletzt wird. Die Pausen signalisieren im allgemeinen das Wortsuchen, den Holper. Die Rolle der strukturellen Gliederung-Verknüpfung der "Sätze" fällt oft auf die Intonation, bzw. die Betonung.

4. Eine wichtige andere Variante der Realisierung der gesprochenen, verklingenden parole ist das Vorlesen, oder Reproduktion oder Interpretation.

Das Wesentliche dabei ist, daß die verklingende parole in jedem Falle auf einer früheren schriftlichen Kommunikation mit gebundenem Text fußt. Zuerst findet die sprachliche Formung der Gedanken (Redigierung, Abfassen) den Gesetzmäßigkeiten der schriftlichkeit entsprechend statt, dann begegnet der Rezipient der verklingenden (eingelernten oder eingelernt interprätierten) Variante des Textes, wobei der schriftliche Kode nicht immer vom Verfasser, sondern sehr oft vom Nachrichtensprecher, Vortragskünstler, Schauspieler zum akustischen Kode umkodiert wird. Eine solche verklingende parole kann auch in der direkten interpersonellen Kommunikation erscheinen, z.B. bei der Verlesung eines Vortrages oder einer Festrede, beim Vortragen eines lyrischen, prosaischen, eventuell Dramawerkes im Vortragssaal oder Theater), kann aber auch in der Form der indirekten, übertragenen oder gemischten Kommunikation erscheinen.

Die so verklingende Rese müssen wir deshalb von der spontanen lebendigen Rede trennen, weil einerseits die Art und Weise der Abfassung den Regeln der schriftlichen Satzkonstruktion entspricht.
Infolgedessen ist auch die Klangwelt (Intonation, Rhytmus) anders: Sie entspricht der latenten Melodienwelt des geschriebenen Textes. Dies bedeutet, daß der Tonfall des verlesenen oder interprätierten Textes sich in gehobener Tonlage in engerem Tonumfang, jedoch mit in größeren Bögen "ausgesungenen" Melodien bewegt und der "Vorhersagewert" der Melodieführung kraftvoller ist als bei der lebendigen Rede. Die Wort- und Satzkonstruktionen werden durch die Intonation zusammengefasst, die Wortkonstruktionen mit zwei Gesichtern[5] ausgelegt und abgesondert. Gleichzeitig ergeben sie eine klare Vorhersage über Aufbau, Länge, Abschließung der Satzstruktur, über Aufbau, Länge, Abschließung der Satzstruktur, über die Anknüpfung des nächsten Satzes. Die Klangwelt der Verlesung charakterisiert auch ein ausgeglichenerer Rhytmus und Tempo (weniger pulsirend als die spontane lebendige Rede). Die Gliederung des Textes mit Pausen oder Intonation und Tonfall fällt fast immer auf die Grenzen der Konstruktionen (Wortkonstruktionen oder Gliedsätze). Die solche, auf geschriebenem Text fußende Reproduktion kennzeichnet auch, daß die die Attitüde der Rede oder die Veränderung der Attitüden signalisierenden Veränderungen (Lautfärbung, Rhytmuswechsel, Lagenwechsel usw.) immer ausgeklügelter, bewußt und deshalb fast nie spontan wirken könne. Sie verraten sich immer.

Das Interprätieren-Verlesen wird in seinen geschilderten allgemeinen Merkmalen durch jene Redeabsichten weiter gefärbt, mit denen der Sprecher die Lautstärke des Textes beeinflußt: Das Bedürfnis einer reellen Mitteilung, die Darstellung einer vollkommenen Indentifizierung, der Ausdruck des "Außenstehens" oder der Abgrenzung, dann das Zitat, die Annäherung an

die spontane Rede, die Erläuterung, die Versinnlichung der Gefühle usw. Mit der Ausnahme einiger Absichten können sie die Vorstellung der spontanen Rede und der vollständigen Identifizierung mit dem Inhalt nie erwecken. Da das uns bereits in die Welt der Vortragskunst führt, besteht da keine Möglichkeit die akustischen Merkmale zu erörtern.

Die wichtigste Bibliographie

DEME László: A kiejtés törvényeinek tanítása és tanulmányozása. Nyr. 94(1970): 270—80.
DEME László: A bemondói munka mondatfonetikai kérdéseiről. In: A rádióbemondó beszéde. (Szerk.: Wacha Imre) MRT TK 1973. 71—101.
DEME László: Grammatikai képlet és akusztikai képlet kapcsolatához. Magyar Fonetikai Füzetek 3. sz. 1979. 7—13.
DEZSÉRY Judit—Terestyéni Tamás: Élő szöveg - stúdió-szöveg. In: ÁNyT XI (1976): 51—77.
ELEKFI László—Wacha Imre: Textliche und intonatorische Faktoren der Sprechwirkung. In: Sprechwirkung. Martin-Luther-Universität Wissenschaftliche Beitrage, Halle, XXIV(1976): 67—74.
FÁBRICZ Károly: A beszélt nyelvi szövegalkotás kérdéséhez. In: Beszélt nyelvi tanulmányok. (Szerk.: Kontra Miklós). Linguistica. Series A. Studia et dissertationes, 1. MTA Nyelvtudományi Intézet, 1988. 76—89.
FÓNAGY Iván—MAGDICS Klára: A magyar beszéd dallama. AkK. 1967.
LÁSZLÓ Zsigmond: Ritmus és dallam. Zeneműkiadó, 1961.
HETZRON Róbert: Ízelítő a magyar tonoszintaxisból. In: A magyar nyelv grammatikája. NytudÉrt. 104. sz. 1980. 389—98.
WACHA Imre: Az elhangzó beszéd szövegfonetikai eszközeinek rendszere és összefüggései. NyK. 75(1973): 77—103. - System und Zusammenhänge der textphonetischen Ausdrucksmittel. ALH. 25 25(1975): 39—75.
WACHA Imre: A bemondói beszéd akusztikus stílusának gondjairól. In: A rádióbemondó beszéde (Szerk.: Wacha Imre). MRT TK 1973. 103—68.
WACHA Imre: Az elhangzó beszéd főbb akusztikus stíluskategóriáiról. ÁNyT. X(1974): 302—16.
WACHA Imre: A tételhangsúlyról. Nyr. 104 (1980): 85—89.
WACHA Imre: Beszéd, szituáció, szöveg és hangzás együttese a rádióban és a televízióban. In: Nyelvészet és tömegkommunikáció (Szerk.: Grétsy László). Tömegkommunikációs Kutatóközpont, 1985. I.
WACHA Imre: Élő nyelvi (spontán) szövegek megnyilatkozásainak (szintaktika) vizsgálati szempontjaihoz (A gazdagréti kábeltelevízió élő nyelvi felvételei alapján). In: Beszélt nyelvi tanulmányok (Szerk.: Kontra Miklós). Linguistica. Series A. Studia et dissertationes 1. MTA Nyelvtudományi Intézet, 1988. 102—158.

ON THE WORD PROSODY OF THE BALTIC-FINNIC LANGUAGES

Kalevi WIIK
Turku, Finland

In Northern Europe there still are half a dozen languages that are spoken by some remote cousins of the Hungarians. The languages are Finnish, Carelian, Vepsian, Estonian, Votian, and Livonian. The intonational and durational patterns of the dialects and standard languages have been studied by Finnish and Estonian phoneticians for over a century. It is my purpose to give a concise summary of some of the results.

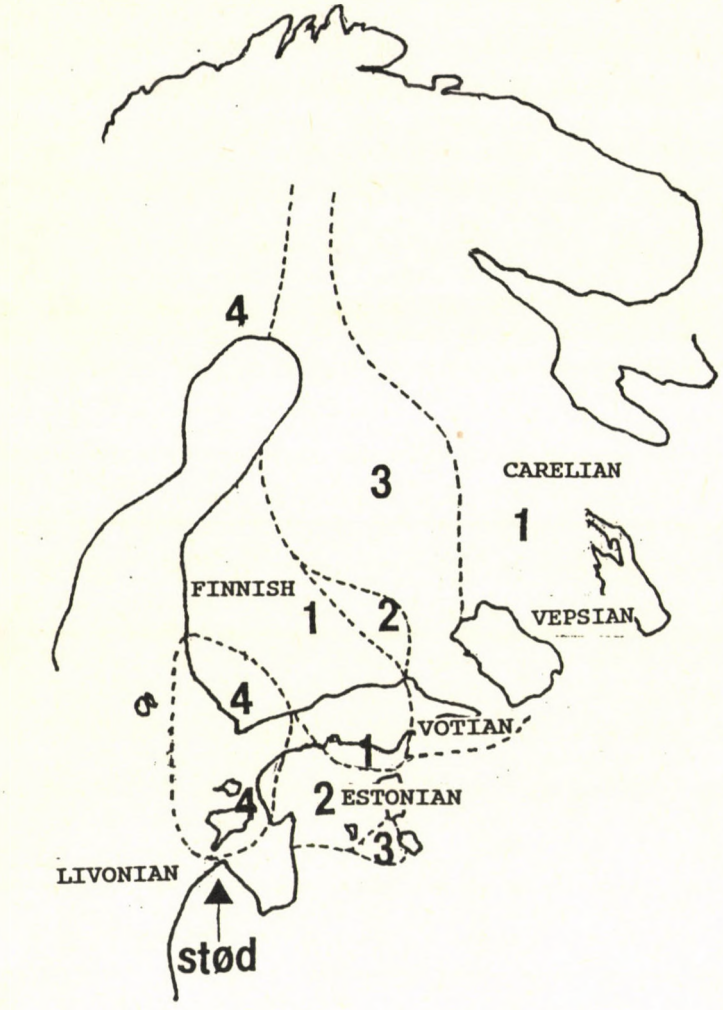
Perhaps the most striking prosodic feature common to all of these languages is the isochrony of foot. This isochrony has several reflections. So for example, the duration of the second syllable vowel depends on the duration of the first syllable in such a way that there is an inverse relationship between the two: the longer the first syllable, the shorter the second syllable vowel; the V_2 of the word type *mana* is always longer than the V_2 of the word types *manna* and *maana*. This type of foot isochrony probably exixted in Proto-Germanic and was taken over by the Baltic-Finns about 2000-3000 years ago.

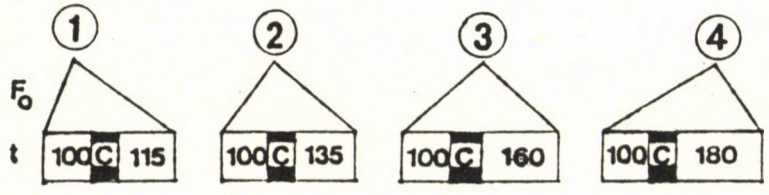
The individual dialects of the languages show systematic differences in their intonational and durational patterns. The word type *mana* shows the following regularities (see the Chart): In Dialect Type 1 (Central Finnish, Carelian, Vepsian, and Nort-Eastern Estonian) the FØ peak is early (during the first vowel segment of the word) and the V_2 is relatively short (about 115% of the V_1). In Type 2 (Central and Northern Estonian and a small South-Eastern area of Finnish) the FØ peak is later and the duration of V_2 is about 135% of that of V_1. In Dialect Type 3 (Eastern Finnish and South-Eastern Estonian) the Fø peak is still later and the duration of the V_2 is about 160% of that of the V_1. In Dialect Type 4 (South-Western and Nort-Western Finnish and the Westernmost type of Estonian) the FØ peak is very late (often on the V_2) and the duration of the V_2 is about 180% of that of the V_1. In Dialect Type 4 (and to a lesser degree in Dialect Type 3) there are two allotones: in the word type *mana* the FØ peak is considerably later than in the word types *manna* and *maana*. It is likely that the two allotones originally come from Swedish (which is a tone language).

In addition, Estonian and Livonian have "gradation of syllables". In Estonian, the first syllable is extra long (3rd quantity) if the second syllable was open (*saùna-ta, saùna-han*), but it is only long (2nd quantity) if the se-

cond syllable was closed (*saunan, saunat*). This is the origin of the three quantity system of Estonian (1st quantity is represented by the word type *kala*.) The intonation pattern is different in the three quantities in such a way that the FØ peak is earliest in the 3rd quantity and latest in the 1st quantity. - In Livonian, there is a similar opposition in the first syllables, but the conditioning factor is different: the first syllable is "strong" if the second syllable has lost its vowel or a contraction of the second and third syllables has taken place (*aɪg* < *halko*); if the second syllable vowel is still there the first syllable is "weak" (e.g. *rānda* < *ranta*). (In the strong syllable the segment following the syllable nucleus is lengthened, while in the weak syllable it is the nucleus itself that is lengthened.)

Livonian also has a stød. Phonetically it is very similar to the Danish stød; its acoustic manifestation is usually a dip in FØ and intensity + glottalization. The Livonian stød was originally a syllable boundary that was situated in the most unmarked position, in front of a CV sequence (*va.lo* and *ka.la*). Later the segment structure was changed and the syllable boundary was no more in the most natural position (*va.l* and *ka.lla*). It remained as "a relic of a syllable boundary". Today it is perceived to be a stød. Also in more general terms, I believe that what is often meant by the stød is a phonetic phenomenon that still has the phonetic qualities of a syllable boundary but which no more functions as a syllable boundary.


4
3
CARELIAN
1
FINNISH
1
2
VEPSIAN
4
VOTIAN
1
4
2
ESTONIAN
3
LIVONIAN
stød


1
2
3
4
F_o
t
100 C 115
100 C 135
100 C 160
100 C 180

AUTHOR INDEX

AN ARTICULATORY STUDY OF "VOICING" IN ITALIAN BY MEANS OF DYNAMIC PALATOGRAPHY

Edda Farnetani
Centro di Fonetica del C.N.R., Padova, Italy

Introduction

Previous acoustic studies on Italian stop consonants indicate that voiceless stops can be classified phonetically as voiceless unaspirated (due to their short positive VOT values) and voiced stops, always characterized by negative VOT values, as fully voiced (1). Intraoral pressure and airflow measurements on one subject (2), besides confirming the acoustic results, suggest that voicing is maintained during oral closure by an active expansion of the oral cavity: this was inferred in bilabial word-initial voiced stops from the presence of an interval of negative pressure during the closure just before onset of vocal fold vibration, and in intervocalic geminated voiced stops by a decrease in intraoral pressure with a concomitant increase in the amplitude of the periodic signal. Moreover, in vowels before voiceless fricatives, but not in those before voiceless stops, a remarkable increase of oral flow and a decrease in the amplitude of the periodic signal was detected just before the increase of intraoral pressure, indicating that the onset of the laryngeal abduction gesture occurs before constriction in fricatives, but not before implosion in stops. The overall aerodynamic data suggest a different glottal behavior for voiceless fricatives vs stops, i.e. an ample abduction/adduction gesture for fricatives vs a quite limited gesture for stops, which seems to start and be concluded within the interval of oral closure, if it occurs at all - it is well known that devoicing can occur with adducted glottis as a result of equalization of subglottal and supraglottal pressure (3).
The present electropalatographic (EPG) study explores the spatiotemporal patterns of the tongue-to-palate contact for lingual stops and fricatives in an attempt to answer the following: are there any differences between voiceless and voiced stops in terms of oral gestures and configurations, and, if so, do they suggest a substantial contribution of the supraglottal gestures to the realization of the voiced/voiceless distinction? Do the voiced vs voiceless linguopalatal configurations differ in stops and fricatives? If the timing of onset/offset of the laryngeal gesture can be inferred from the voice offset/onset times relative to the articulatory configurations, do the data suggest that devoicing is brought about in a similar fashion for stops and fricatives, as seems to occur in English (4)? In this investigation consonant clusters as well as word-initial and word medial intervocalic consonants were studied. However, in the present report I shall focus, on initial and intervocalic consonants due to space limitations.

Procedure

Synchronous acoustic and EPG recordings were made of 5 readings, by 3 speakers of standard Italian, of a list of paroxytone words with the stops /t-d/, /k-g/ in word initial and in intervocalic positions and with the [s-z] allophones of the alveolar fricative in intervocalic position, according to the phonotactic rules of Italian. The vocalic contexts were /i/ and /a/. We used the EPG system of the Reading University (5) (scanning frequency= 10ms). An example of the EPG printout is displayed in Fig.1. Closures and constrictions have been defined palatographically, and checked when necessary with the intensity

curves and/or the duplex oscillograms on mingograms. The linguopalatal configuration contact has been represented for each consonant by the relative frequency (in percent) of electrode activation during the interval of closure for stops and of constriction for fricatives. The following temporal parameters were measured: voice offset and voice onset time for voiceless stops (Fig.1: B-B′, C-C′), duration of voicing for voiced Cs, duration of closure for stops (Fig.1: B-C) and of constriction for fricatives, duration of the total V-to-V gesture (Fig.1: A-D).

Results

Temporal intervals. For the voiceless stops, average VOT values are 25 ms for /t/ and 50 ms for /k/, the vocalic context also affects VOT, which is higher in the context of /i/ than in the context of /a/ (/ta/=15ms vs /ti/=34ms). Voice offset occurs after implosion in 76% of total occurrences, at the instant of implosion in 20% and before implosion in only 4% of the cases. Voiced stops, in both initial and intervocalic position always have negative VOT values. Short intervals of devoicing were observed most often in intervocalic position and in the context of /i/. As for voiced fricatives there are systematic voice breaks around the maximum constriction points (45 ms for S1, 83 ms for S2 and 78 ms for S3); these breaks are always shorter than the duration of the constriction and about half the duration of the aperiodic interval that characterizes the voiceless fricatives. Closure durations, constriction durations and V-to-V intervals all tend to be longer in voiceless than in voiced Cs; the trend is stronger in stops than in fricatives. See Fig. 2 for the mean durational values in S1. For stops, the average respective durations of voiceless and voiced closures are 119 and 71 ms for S1, 115 and 66 ms for S2, 113 and 73 ms for S3.

Linguopalatal contact patterns. The palatographic frames in Fig.3 represent the difference in linguopalatal contact between voiced and voiceless Cs for S1. The configurations are based on comparisons of the mean percent frequency of contact during closure or constriction. The filled surfaces are the common areas with 100% frequency of contact; the empty surfaces are the common areas of no contact. The striped zones indicate that contact occurs for both voiced and voiceless but does not reach 100% frequency (in either or in both) and the numbers represent the difference in percent frequency (positive figures= higher frequency for voiceless); the numbers in the white areas indicate that contacts occurred only for one of the two categories and represent the frequency of contact (positive= contact for voiceless). It can be seen that the voiced/voiceless relation is strikingly different for stops and fricatives: voiceless stops have a larger field of contact than their voiced counterparts, while for fricatives the reverse is true. The data for the other two subjects confirm this general trend. The largest intersubject differences are for the stop consonants: in the context of /i/ in S2 and S3, the difference in the contact area between voiced and voiceless tends to be smaller than in S1, and in some cases there is a perfect overlap of the covered surface.

Discussion

The picture that emerges from the overall temporal and configurational data is of a difference in patterns of spatiotemporal control for the oral articulations of the voiced and voiceless consonants. For stops, the less extended contact area in most of the voiced closures than in the voiceless, suggests a less constricted oral volume favouring the maintenance of fold vibration; at the same time their very short

durations automatically prevent long intervals of aperiodicity, in relation to the total duration of the closure, when devoicing happens to occur. So, both temporal and spatial features contribute to the voiced/voiceless distinction in a precise way that can be interpreted as a direct contribution to the maintenance of vocal fold vibration in voiced stops. For fricatives the voiceless combine less constriction with longer duration, while the voiced counterparts combine greater constriction with a shorter duration. These features suggest that for the voiced fricatives an important aim of the speakers was not so much to maintain fold vibration, as to produce a friction noise of the appropriate quality, by means of a more constricted oral volume, since the velocity of the transglottal flow is rather low with the adducted glottis. (Such a maneuver is unnecessary for voiceless fricatives, for which the VF are assumed to be abducted.) A more constricted oral volume, however, favours voice breaks, which in fact most often characterize portions of the constriction intervals of voiced fricatives. On the other hand, the shorter durations of the voiced fricatives in relation to the voiceless would suggest that speakers also aimed at keeping the aperiodic interval rather short presumably to enhance the perceptual difference between the intervals of aperiodicity in the voiced vs voiceless fricatives. A direct comparison between stops and fricatives indicates that they differ in the realization of the voiced/voiceless distinction at both glottaland supraglottal levels. The oral articulation for stops and fricatives indicates that in the former, both temporal and spatial commands contribute to enhance the voicing distinction, while in the latter there is no such association. Moreover, from the differences between (voiceless) stops and fricatives both in voice offset and onset times in relation to the oral configuration and in the total durations of the aperiodic intervals, we can infer that the articulation of stops and fricatives differ also at the glottal level. This confirms previous aerodynamic data indicating that voiceless fricatives are produced with a longer anda more extended abduction/adduction gesture than voiceless stops.

Acknowledgments: The research was supported in part by a grant from NATO

References

1) Vagges, K., Ferrero, F. Caldognetto-Magno, E., Lavagnoli C.: Some acoustic characteristics of Italian consonants. J. of Italian Linguistics 3. 1978, 69-85.
2) Farnetani, E.: Profilo aerodinamico di alcune consonanti italiane. Acta Phoniatrica Latina, 1. 1985, 33-45.
3) Rothenberg, M. : The breath-stream dynamics of simple-released-plosive production. Biblioteca Phonetica, 6. 1968.
4) Weismer, G.: Control of the voicing distinction for intervocalic stops and fricatives: some data and theoretical considerations. J. of Phonetics, 8, 1980, 427-438.
5) Marchal, A.: La Palatographie. Editions du CNRS, Marseille, 1988.

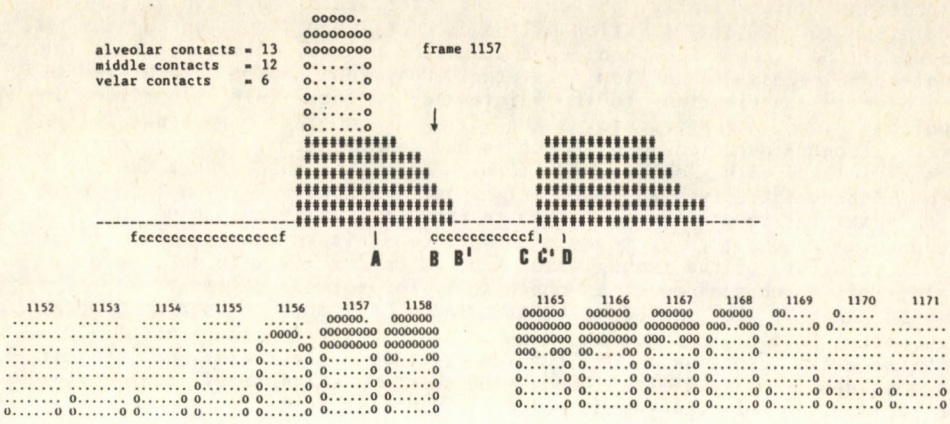


Fig. 1 Pattern of acoustic intensity and selected EPG frames for the production of /'tata/ (S2).

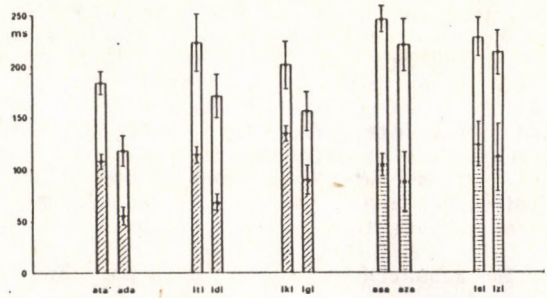


Fig. 2 Vowel-to-vowel mean durations divided into closures/constrictions (striped) and transitional phases (unfilled) (S1).

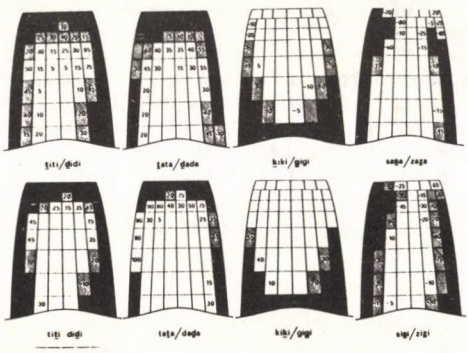


Fig. 3 Differences in linguopalatal contact area between voiceless and voiced stops and fricatives (S1) (see text for explanation).

Címünk:

MAGYAR FONETIKAI FÜZETEK

A Magyar Tudományos Akadémia
Nyelvtudományi Intézete
Fonetikai Osztály

Budapest, I., Szentháromság u. 2.
1014

Address for communications:

HUNGARIAN PAPERS IN PHONETICS

Department of Phonetics
Linguistics Institute, HAS

Szentháromság u. 2.
Budapest
H-1014